Northwestern University Society for the Theory of Ethics and Politics

11th Annual Conference March 30-April 1, 2017

John Evans Alumni Center-1800 Sheridan Rd., Evanston

Keynote Addresses:

Other People

Kieran Setiya, MIT

Objective and Subjective Standards of Reasonableness in the Law of Self-Defense

Marcia Baron, Indiana University

A SPECIAL THANK YOU TO THE GRADUATE SCHOOL AND THE WEINBERG COLEGE OF ARTS AND SCIENCES FOR THEIR GENEROSITY

Table of Contents

Northwestern University Society for the Theory of Ethics and Politics

11th Annual Conference
March 30 – April 1, 2017
John Evans Alumni Center

Thursday, March 30th 2017

Morning Session

9:00-10:25
You Can't Move without Being Moved: On the Moral Significance of The Human Capacity for Feeling
Anastasia Artemyev Berg (University of Chicago)
Comments: Carmen De Schryver (Northwestern)

10:35-12:00
Belief, Intention, and Deliberation
Gregory Antill (Claremont McKenna College)
Comments: Grant Rozeboom (St. Norbert College)

Lunch

Afternoon Session

2:15-3:40.
Contingency and Integrity
Joshua Tignor (Syracuse University)
Comments: Hao Liang (Northwestern)

3:50-5:15
Love, Reason, and the Highest Good
David Sussman (University of Illinois, Urbana-Champaign)
Comments: Paul Schofield (Bates College)

Dinner

Friday, March 31st 2017

Morning Session

9:00-10:25
Hoping for Peace
Lee-Ann Chae (University of Pennsylvania)
Comments: Per-Erik Milam (University of Gothenburg)

10:35-12:00
Biographical Identity and Retrospective Attitudes
Camil Golub (New York University)
Comments: Michael Schwarz (Northwestern)

Lunch

Afternoon Session

2:15-3:40
Intelligibility and the Guise of the Good
Paul Boswell (Université de Montréal)
Comments: Nandi Theunissen (Johns Hopkins University)

3:50-5:45.
Keynote Address:
Sexual Consent, Reasonable Mistakes, and the Case of Anna Stubblefield
Marcia Baron (Indiana University)
Comments: Louis-Phillipe Hodgson (York University)

Reception – Everyone is invited

Saturday, April 1st 2017

Morning Session

10:35-12:00
Do I Have To Be Coherent To Be Reasonable?
Alex Schaefer and Wes Siscoe (University of Arizona)
Comments: Henry Andrews (Northwestern)

Lunch

Afternoon Session

2:15-3:40.
Answerability Without Blame?
Andrea C. Westlund (Univeristy of Wisconsin Milwaukee)
Comments: Mihailis Diamantis (University of Iowa)

3:50-5:45
Keynote Address:
Other People
Kieran Setiya (MIT)
Comments: Jennifer Lockhart (Auburn University)

Dinner

Acknowledgements

Conference Organizers:

Kyla Ebels-Duggan, Richard Kraut, Stephen White, and Abby Bruxvoort

For organizational and administrative assistance, special thanks to Morganna Lambeth, Josh Kissel, Will Cochran, Hao Liang, Whitney Lily, Blaze Marpet, Andy Hull, Crystal Foster, Jasmine Hatten, and Ozge Hemmat.

You Can't Move without Being *Moved*
On the Moral Significance of The Human Capacity for Feeling

Anastasia Artemyev Berg

Abstract: Kant's account of moral respect is supposed to answer the question of how Kantian moral judgments are *motivating*. It poses however an apparently insoluble exegetical challenge: on the one hand, moral action is supposed to be autonomous and as such independent of feelings. On the other hand, the "feeling of moral respect" is necessary for moral action. Interpreters are divided between "intellectualists" who jettison Kant's account of the involvement of feeling in moral motivation (rendering mysterious how moral judgments motivate at all), and "affectivists" who claim that respect is "non-pathological" and can therefore safely motivate moral action. I demonstrate that affectivists fail to secure a characterization of "non-pathological" feeling that's adequate to account for Kant's incisive critique against the involvement of feeling in *moral* motivation. I claim that the distinction between pathological and non-pathological feeling should instead be understood by reference to Kant's distinction between the lower and higher faculty of desire: between the kind of desire that subrational animals have and *our* rational faculty of desire, the will. If our faculty of desire is unique so must be our faculty of feeling. This faculty of feeling Kant shows us must be understood as a form of self-consciousness, which (1) constitutes and reveals the agent to be practically rational and (2) is the basis of all particular feelings.

§1. Introduction

Kant grounded his moral system in the thought that to do what is right is to do what reason prescribes, whether or not it serves your personal interests, whether or not you happen to feel like it. According to the dominant reception of Kant, advanced by his sympathizers and critics alike, this principle and the system to which it gives rise are grounded in an irreducible dualism between, on the one hand, our rationality and freedom and, on the other, our animality and feeling.

According to this traditional interpretation, Kant does violence to some of the most fundamental aspects of human experience. First, in our common ways of self-understanding we do not encounter an in-principle unbridgeable gap between our emotional states and our rational mental activity. Of course we find ourselves often questioning whether a particular emotion is reasonable or justified, but it is precisely this act of questioning which presupposes the possibility of feelings that are reflective of our rational commitments. (We consider it *reasonable* and indeed *appropriate* to feel indignation in response to moral injustice or to feel regret at the realization of wrongdoing.)

Second, a dualism of rationality and feeling threatens the internal coherence of the Kantian account itself. Specifically, it would renders it difficult for Kant to provide a convincing account of how we are *motivated* by the moral demands he labors to articulate. To be capable of motivation by a moral demand is to be capable of doing

something just because one has recognized that doing so is the right thing to do. In other words, it is to *be moved* to do the right thing in light of the recognition of an action's goodness. This requires our capacity *to feel* to be responsive to the claims of reason, as Kant himself insists: a feeling (namely, moral respect) is necessary for the performance of moral action. A picture where our emotional lives are divorced from our lives as free and rational cannot support the idea that feeling could come to reliably manifest the demands of reason, and since it renders unintelligible the idea that a concern with the right thing to do could ever come to manifest itself affectively, it can hardly be understood to move us to act.

Scholars sympathetic to Kant's account of morality, which promises to secure objective universal moral claims, have tried to resist these implications by rehabilitating Kant's treatment of human feeling. In particular, commentators have attended to the feeling of moral respect—which Kant defines as the necessary effect of reason on sensibility in the determination to moral action—as a key to understanding the relation of reason and feeling in Kant generally and securing the role of feeling in moral motivation specifically. Nevertheless, a coherent reading of the role of feeling in moral motivation—and with it Kant's profound insight into the distinctive role feeling plays in the life of a rational animal—have so far eluded commentators.

§2. THE INTERPRETATIVE PROBLEM

Kant's account of the feeling of moral respect poses an apparently insoluble exegetical and philosophical puzzle.[1] On the one hand, moral action is supposed to be autonomous—in acting from the moral law I set my own ends in accordance to principle—and as such must be independent of any external determination, and thus independent of feelings. On the other hand, Kant claims that the "feeling of moral respect" is a necessary moment of acting from the moral law: "[i]mmediate determination of the will by means of the law and consciousness of this is called respect" (*G* 4:400n).

Feelings, according to the conventional understanding of Kant, are reflective of our contingent, non-rational animal nature and empirical habituation, we can only come to know them empirically, observe them in experience. Therefore they cannot determine or influence the will to perform an action that has moral worth.[2]

§3. THE INTERPRETATIVE DEBATE

Parties to the interpretative debate customarily share the following assumptions:

1. "Respect" is necessary for the performance of morally worthy, i.e., free action.

[1] So much so that Robert Wolff famously charged that "the introduction of the emotion of reverence [respect] is contradictory to the entire thrust of Kant's argument."

[2] Onora O'Neill provides a clear statement of the worry about attributing any role, motivating or otherwise, to the moral feeling of respect or reverence: "To act 'out of reverence [respect] for the law,' is not to act with any peculiar ***feeling*** of reverence or awe. [...] Pathology, as Kant would have it—psychology, as we would say—is irrelevant to the moral worth of acts." See Onora O'Neill (Nell), *Acting on Principle* (New York, 1974), 111.

2. There is a distinction between two aspects of respect: a "purely intellectual recognition of the supreme authority of the moral law" (defined negatively by having nothing to do with feeling) and "a peculiar moral feeling of respect for law."[3]
3. When Kant speaks of feeling, moral or otherwise, he is speaking of various effects on a single "faculty:" a sensible, receptive capacity for feeling whose exercises we recognize as pleasure and pain.

Two camps emerge in the interpretative debate, affectivists and intellectualists. The distinction between them, following Richard McCarty, turns on the question, "whether the affective component of respect plays any role in the mechanism of moral motivation."[4] That is, does *feeling* play a role in the determination of morally worthy action?

"Intellectualists" deny any role to the *feeling* aspect in the determination to moral action[5]. Andrews Reath, for example, writes, "it is the practical aspect [of respect] that is active in motivating moral conduct, while the affective side, or *feeling* of respect, is its effect on certain sensible tendencies."[6] In this picture, a "practical [intellectual] aspect of respect" is wholly responsible for acting as one should, and no feeling is required to explain why the agent acted as she did when she acted from respect for the moral law. Feelings merely accompany this moral activity as a side-effect.

The challenge to the intellectualists is exegetical. Whereas the affectivist can easily point out to many passages that suggest that the "feeling aspect" of respect is necessary to motivate moral conduct, intellectualists have tried to dismiss these passages (for example Kant's pronouncements reported by his students and published in the *Lectures on Ethics* (MPC, AA 27:1428) or even the *Groundwork*, 4:400-401) as expressions of an early, abandoned view.[7] Yet this intellectualist suggestion becomes impossible to sustain in the face of passages from Kant's later writings (for example,

---

[3] Richard McCarty, "Kantian Moral Motivation" (1993), 421.
[4] Ibid, 430.
[5] The paradigmatic proponent of the intellectualist position is Andrews Reath, "Kant's Theory of Moral Sensibility," (1989). Jens Timmermann holds the cultivation of moral feelings is an indirect duty concerning the acquisition of instrumental means for implementing moral ends. Moral feelings are thus not valuable in themselves. They are like acquiring prosperity, which makes it easier to act morally. Timmermann, "Kant on Conscience,"Indirect" Duty, and Moral Error" (2006), 298-302. This view is also defended by Henry Allison, *Idealism and freedom: essay on Kant's theoretical and practical philosophy,* (1996), 123, and Pablo Muchnik, "The Heart as Locus of Moral Struggle in Religion," in *Kant on Emotion and Value* ed. A. Cohen, (2014), 233-4. Marcia Baron claims emotions cannot have a motivational function because of impurity concerns and argues that feelings only have a supportive epistemic function: they "*help* to direct ... our attention to the needs of particular others and to ways we might help" *Kantian Ethics Almost without Apology* (1995), 220. Nancy Sherman claims moral feelings are not necessary for moral action but are a "layer of character that can ... best support moral motivation," they "positively promote ... our duty motive." In addition to this subsidiary motivational role, there is also an epistemic function of moral emotions: they are "modes of attention that help us to track what is morally salient ... in our circumstances." Nancy Sherman, "Kantian Virtue" in *Making a Necessity of Virtue*, (1997: 144-146). A similar view is held by Anne Baxley, *Kant's Theory of Virtue: The Value of Autocracy* (2010:124, 135, 136, 145, 164). See also, Onora O'Neill, *Acting on Principle* (1974). Paul Guyer , "*Kant and the experience of freedom,*" (1993). Stephen Engstrom, "The *Triebfeder* of Pure Practical Reason," (2010).
[6] Andrews Reath, "Kant's Theory of Moral Sensibility," (1989), 287.
[7] See, O'Neill, (1974) and Robert Wolff, *The Autonomy of Reason*, (1973), 83.

KpV 5:75, 5:79). In dismissing these passages, we will see, the intellectualists do not only ignore the letter of Kant's text, but miss Kant's deep insight into the distinctive role feeling plays in the lives of free, rational beings.-

"Affectivists," in an attempt to secure for moral respect a unique status that would legitimize its role in moral motivation, appeal to Kant's distinction between "pathological" and "non-pathological" feeling and claim that only pathological feeling poses a threat to freedom, while non-pathological feeling, the kind of feeling that is operative in "moral respect," can happily play "a motivating role" in moral action.

That is, affectivists claim that the affect produced when the moral law determines the will, i.e., the feeling aspect of respect, is responsible for the *motivation* to moral action.[8] The affectivist, on McCarty's characterization, "need not deny that Kantian moral motivation initially arises from an intellectual recognition of the moral law. Contrary to intellectualists, however, they maintain that it ***also*** depends on a peculiar moral feeling of respect, for the law, one consequent to the initial recognition or moral judgment the intellectualists emphasize exclusively."[9] The affectivist account of moral respect thus identifies two necessary steps for the performance of morally worthy action: intellectual recognition followed by moral feeling. It is the latter that explains why the agent, having "recognized" the moral law, was moved to act.

What of the intellectualist's worry of heteronomy? McCarty, on behalf of affectivists, identifies the implicit "erroneous" assumption that gives rise to this worry as the "classification of all feelings as pathological."[10] It is because respect for the moral law must be capable of motivating action independently of the typical, empirical, contingent, or as Kant calls them, "pathological" motivational resources of human agency, McCarty points out, that commentators assume that "respect for the moral law motivates independently of any feeling or affections whatsoever".[11] Embracing the letter of the text, namely Kant's insistence that respect in its role in motivation is a feeling, seems therefore as easy as denying that all feeling is pathological and acknowledging that, specifically, respect is not.

At this point, the affectivist position faces two challenges. (1) She must provide a characterization of non-pathological feeling that withstands Kant's critique of the involvement of feeling in moral motivation; (2) The affectivist faces both philosophical

---

[8] A paradigmatic proponent is Richard McCarty, "Kantian Moral Motivation" (1993). In later work McCarty conceives of moral emotion as a "psychologically forceful incentive" that aids agents to do what duty requires by outweighing other motives. Moral feelings are pleasures and displeasures, which allows us to say that "the maxim incorporating the motivationally stronger incentive" prevails in cases of conflict. McCarty, *Kant's Theory of Action* (2009), 167, 182. Christine Korsgaard claims that our capacity for respect names the rational will's capacity to provide "not only the ground of choice but ***also*** the incentive to act in accordance with that ground" and goes on to demonstrate how the moral law could provide painful and pleasurable feelings that could serve as this incentive. "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action," *The Constitution of Agency* (2008), 187 (emphasis mine). See also, Owen Ware, "Kant on Moral Sensibility and Moral Motivation" (2014), Larry Herrera "Kant on the Moral Triebfeder" (2000), Ido Geiger "Rational Feelings and Moral Agency," (2011), Jeanine Grenberg, "Making Sense of the Relationship of Reason and Sensibility in Kant's Ethics" (2011).

[9] McCarty, 423.

[10] McCarty, 424.

[11] *Ibid*.

and textual challenges to her reliance on a *distinction* between intellectual recognition and an affective response, which is required for her interpretation. Identifying and addressing these two challenges will lead us towards an alternative interpretation.

§5. First Challenge: Characterizing Non-Pathological Feeling

Affectivists must characterize non-pathological feeling in such a way that it does not collapse into empirical and contingent feeling. Recognizing this necessity, affectivists point to moral respect's allegedly intellectual and non-natural (i.e., not empirically cognizable) cause. Larry Herrera, for example, writes, "we all know that for Kant there is a fundamental difference between respect and sensuous feelings. The former is effected by pure reason alone; the latter, sensuously so. Thus, although all feeling is sensuous, not all feeling is pathological."[12] In this account, respect is not pathological *by virtue of the source of the cause that effects it*. Specifically, in non-pathological feeling the representation which affects our faculty of feeling and induces the feeling of moral respect is a representation of the moral law, it is a not representation of an object of the senses, but *a representation of reason.*

However, this interpretive suggestion, broadly adopted and unchallenged, faces a major exegetical challenge that reveals in turn a deep philosophical problem. Kant vehemently objects to the idea that the source of the representation can be used to distinguish between feelings as far as the moral worth of the actions they motivate is concerned. Early and prominently in the second *Critique* (Part I, §3) Kant explicitly denies that the origin of a representation in reason could secure for it a legitimate motivating role in the determination to moral action.

> However dissimilar representations of objects may be—they may be representations of the understanding *or even of reason,* in contrast to representations of sense—*the feeling of pleasure by which alone they properly constitute the determining ground of the will* (the agreeableness, the gratification expected from the object, which impels activity to produce it) *is nevertheless of one and the same kind not only insofar as it can always be cognized only empirically but also insofar as it affects one and the same vital force that is manifested in the faculty of desire*, and in this respect can differ only in degree from any other determining ground. (*KpV* 5:22-23, emphasis mine).

Here Kant explicitly claims that *no feeling* is to serve as the determining ground of the will, even if it's caused by a representation of reason. Therefore, the affectivists' attempt to distinguish moral feeling from other types of feeling by virtue of the affecting representation's origin in reason is insufficient to secure its legitimate role in motivating moral action. Even if doing the right things would in fact happen to feel good, that promise of pleasure cannot be what motivates the agent to act.

Since in the affectivists' picture the faculty of feeling is conceived as an ordinary, empirical, and sensible faculty of feeling, the only way in which the thought of the moral

[12] Larry Herrera, "Kant on the Moral *Triebfeder*" 2000, 401.

law could induce a feeling would be for our faculty of feeling to be so constituted as to *of itself* respond to a thought of a certain kind, in this case thought "of the moral law," with pleasure and thereby a positive incitement to action. Reason, in this picture, is "responsible" for our recognition of the moral law, and particular moral demands, itself, but it is a feature of our contingent, psychological constitution that this recognition produces a certain feeling. In order to do the right thing we would have to have just the right kind of non-rational sensible faculty, one that would be able to distinguish and privilege, of its own, the source of a representation in reason and thereby respond to inducements by commandments of the moral law.

But even supposing that a sensible faculty could *of itself* respond to representations of reason, in the affectivist's picture the claim that reason is able to motivate us to moral action amounts to the supposition of a certain kind of empirical sensibility. Rejecting this possibility is at the heart of Kant's moral system and his critique of sentimentalism:

> There is here no *antecedent* feeling in the subject that would be attuned to morality: that is impossible, … the incentive of the moral disposition must be free from any sensible condition. (*KpV* 5:75)

The affectivist does precisely what Kant warns against: she assumes an antecedent capacity for feeling in the subject that is "attuned" (i.e., attuned of its own) to the claims of the moral law. In addition to recognizing the law, i.e., recognizing what it would be good to do, in order to *act* on this understanding, the independent cooperation of a sensible non-rational faculty is required. In the case of a human agent doing the right thing, the full explanation of why she acted as she did would have to be: I recognized what the right thing to do is, and, luckily, felt like doing it.

So, while affectivists promisingly direct our attention to Kant's characterization of respect as a non-pathological feeling as the key to resolving the puzzle of feeling's role in moral action, they miss the deeper import of the intellectualist's worry. Whatever the feeling of moral respect is, its role in the determination to moral action cannot be a matter of inducing in a non-rational faculty, by whatever means, a feeling that motivates action. Without recovering an account of non-pathological feeling that is not grounded in a distinction *between the sources of representations that affect the feeling faculty*, the affectivist cannot appeal to the pathological/non-pathological distinction to secure the feeling of moral respect a role in the determination to moral action.

A clue toward an alternative ground for the distinction between pathological and non-pathological feeling will emerge from a consideration of a second challenge to the affectivist, namely, to her reliance on a two-stage account grounded in a distinction between the "intellectual recognition" of the moral law and the "feeling" of moral respect responsible for being motivated by the moral law. A two-stage account, I will go on to argue, misses Kant's radical claim about the unity of the recognition of the law and the feeling of moral respect: on Kant's account, the feeling of moral respect is *nothing but* the mode of recognition of the moral law.

§6. SECOND CHALLENGE: THE DISTINCTION BETWEEN RECOGNITION AND FEELING

As we've seen, the affectivist claims that recognition of the law in the intellectual or practical sense is followed by a feeling that motivates action.

Kant's own characterization of the feeling of moral respect is however subtly and importantly different:

> What I cognize immediately as a law for me I cognize with respect, which signifies merely consciousness of the subordination of my will to a law without the mediation of other influences on my sense. Immediate determination of the will by means of the law and consciousness of this is called respect. (*G* 4:401n)

To recognize the law by the subject of the law as a law for the subject is just "to cognize the law *with respect*." The act of recognition of the law *as a law* for the subject, i.e., the act of recognizing that one is bound by the law, *is itself* the feeling of respect. Thus, respect is the way we recognize the law as a law for us, and this recognition leaves no remainder: nothing further is necessary for the subject to act.

With Kant's characterization of moral respect, Kant is denying the claim that recognition of the law, i.e., rational recognition of demand, and feeling are fundamentally distinct. If you are not moved to act in the manner you've recognized you ought to, nothing is left of the idea of recognition of a demand. A *practical* demand is something you *must* respond to not with "intellectual" assent but *with action*. To recognize a law as issuing in a demand, is to find that demand necessary and binding, i.e., to act in accordance with it and because of it. (In the same way that to recognize a proposition as true is to come to believe it).

If feeling does not play a secondary motivating role in addition to intellectual recognition, in insisting that feeling is necessary for the recognition of the law, Kant can only mean that the feeling of respect is itself the form of recognition of the moral demand: the feeling of moral respect is nothing but the form of recognition of ourselves as bound by moral considerations.

The claim that determination of the will, the recognition of the law as a law for the subject, is itself the *feeling* of respect is reiterated by Kant in the *Critique of Judgment*. The following (sorely overlooked) passage is particularly illuminating. Here, Kant reviews his own treatment of moral respect in the second *Critique* and straightforwardly denies that the feeling of respect is an *effect* of the determination of the will, and explicitly insists that it is instead *identical* with it:

> [I]n the critique of practical reason we actually derived the feeling of respect […] from universal moral concepts a priori. […] there *we could also step beyond the bounds of experience and appeal to a causality that rests on a supersensible property of the subject, namely that of freedom*. […] The state of mind of a will determined by something, however, *is in itself already a feeling of pleasure and is identical with it, thus it does not follow from it as an effect*. (*KU* 5: 221-2, emphasis mine)

Thus, for Kant, the determination of the will by the law, the recognition of the law as a law for the subject, does not effect respect at all, but *is identical* to the feeling of respect for the law.[13]

In order to give content the idea of the identity of feeling and determination of the will, we must turn to Kant's *general* characterization of feeling. We will thereby be in a position to elucidate the precise sense in which respect is a "non-pathological" feeling.

§7. KANT'S GENERAL ACCOUNT OF FEELING

What could Kant's assertion of *identity* between feeling and the state of mind in willing—an exercise of the capacity to act for the sake of ends and in accordance with principles—mean? To answer this question we need to examine and reconstruct Kant's *general* account of feeling.

An analysis of Kant account of feeling will reveal that in speaking of feeling Kant does not have in mind *a single faculty of feeling at all*. Feeling is not another power of the mind: it is not one more ability we have in addition to others, such as the ability to perceive objects around us, to form empirical judgments, or to act according to ends we set for ourselves. It is instead an awareness of how things stand with the moral subject vis-à-vis her own activities and features of her environment.

For Kant, feeling is a kind of sensibility. Sensibility divides into two aspects of susceptibility to representation (1) sensibility as sense, on the basis of which a subject can form judgments of experience and (2) sensibility as feeling:

Feeling, Kant claims, is the "subjective aspect" of our enjoying any representations in general (*MS* 6:211). Feelings Kant claims, "*cannot be explained by themselves at all*" (*KU* 20:231-232, emphasis mine). The "representations" of sensibility as feeling consist only in the awareness of the *relation* of representations (of whatever sort) to the subject. Therefore, "they can be only inadequately explained through the influence that a representation has on the activity of the powers of the mind" (ibid). This influence is "the effect of a representation (that may be either sensible or intellectual) upon a subject," (MS 6:211). This influence upon the subject Kant defines as the "causality of a representation for maintaining (pleasure) and hindering (displeasure) a state of the subject" (5:220). When a representation is causally efficacious with respect to the power of the mind that is the subject's faculty of desire, her capacity to act for the sake of ends, feeling reveals "the causality of a representation for producing its object" (*KU* 20:230-232).

We are now in a position to identify the basic structure of feeling: it is a relation of the subject to a representation, and through it to the object of the representation. Taking pleasure is an awareness of a representation as beneficial for a certain power of the subject and so—since the subject is a unity of various mental powers—it is an

[13] See, *G* 4:401n, 459, *KpV* 5:79, 88, 116-119, *MS* 6:211, 399, *KU* 5:222. Guyer, who holds a similar two-stage reading, curiously claims that there is an "absence of any explicit characterization of [moral respect] in the *Groundwork*" and speaks of the "the introduction of the feeling of respect in the CPrR" (359).

awareness of the subject as benefiting, qua subject of that power. Finding something to be painful is an awareness of a representation as harmful to a certain power of the subject, and so an awareness of the subject, qua subject of that power, as harmed.

Since what benefits or harms a particular power of mind depends on the constitution of each power, feeling does not have its own principles but, in revealing the different powers of the mind as benefited or harmed, reflects and reveals the constituting principles of the different powers of mind themselves—the understanding, imagination, the will. *Because an awareness of its different powers*, feeling is therefore *essentially an awareness of self.*

Most importantly, this means that in this account *there is no distinct faculty of feeling* at all. There is no standalone ability to feel, whose own laws determine what it is the subject will find pleasurable and what painful. Instead feeling reflects and reveals the form and purpose of our various activities.

§8. Feeling and the Will

In the *Second Critique* Kant applies his general understanding of feeling, to the relation between feeling and the essential power of mind that is the faculty of desire—the power to act for the sake of ends.

> The faculty of desire is a being's *faculty to be by means of its representations the cause of the reality of the objects of these representations.* Pleasure is the *representation of the agreement of an object or of an action with the* subjective *conditions of life,* i.e., with the faculty of the *causality of a representation with respect to the reality of its object* (or with respect to the determination of the powers of the subject to action in order to produce the object). (*KpV* 5:8n)

Feeling, in its relation to desire, our capacity to act for the sake of ends, is the representation of agreement or disagreement with, i.e., promotion or hindrance of a subject's exercise of a faculty of desire. However the faculty of desire in rational agents in general and human beings in particular is not merely an animal faculty of desire, but "a specially constituted faculty of desire" (*G* 4:428), "distinct from a mere faculty of desire" by being a "faculty of determining itself to action as an intelligence and hence in accordance with laws of reason independently of natural instincts" (*G* 4:459). By saying that the will is not a "mere" power of desire, Kant is of course not denying that the will is a power of *desire* but claims that our power of desire is a special species of the genus "faculty of desire." The *human*, rational faculty of desire is a faculty of desire, the power to act for the sake of ends, but it is not merely that: it is the capacity to act for the sake of ends *one sets for oneself and in accordance with principle*, therefore in awareness of one's freedom from external determination.

In a being in whom reason is practical, a being with a will and not a power of "mere desire" feeling will reveal how objects and actions benefit or harm the subject's power of willing, i.e., her capacity to act freely.

We are thereby in a position to clarify the precise sense the feeling of moral respect is nothing but the conscious determination of the human will. Since the principle that is the moral law is the form of will, the form of our practical reason,[14].the state of a mind of a will determined by the moral law will necessarily consist in the awareness of this *self-agreement*, and this awareness is nothing but the primary mode of the feeling of moral respect.

We are likewise in a position to fully elucidate the claim that respect is a non-pathological feeling. Moral respect is distinguished as non-pathological feeling not by virtue of the representation (the moral law) which affects a putative "faculty of sensibility" having its source in reason. Instead, non-pathological feeling is the awareness of how things promote or hinder our *rational desire*, or how they promote or hinder *us* as free, rational and embodied beings.

§9. Moral Respect as Practical Self-Consciousness

We saw that the intellectualist is right in insisting that no special feeling motivates the agent to perform morally worthy action, but she is wrong to deny feeling a role altogether. By securing a distinction between pathological and non-pathological feeling that is grounded in the distinction between sensible and rational desire, we are able to avoid the intellectualist's worries of heteronomy and recognize the sense in which feeling is necessary for the performance of morally worthy actions: it is the fundamental awareness of ourselves as moral beings, subject to moral demands, that is necessary for moral action.

Furthermore, Kant claims that the object of the feeling of moral respect is the moral law. This does not, however, mean that the object of this feeling has some sort of fully determinate content on its own; for the moral law, on Kant's account, is nothing but the *form* of our will. This means that the moral law is the *fundamental principle* of our rational, free activities. If the feeling of moral respect has for its object not a particular end or object but this principle, then it is the mode of awareness or the form of one's activity. It is, in other words, the mode of *self*-consciousness that characterizes moral being. Moreover, the moral law is not only the principle of *my* individual activity as free and rational, but also the principle of the activity of every other person—all-pervasive, it is the fundamental principle of everything touched by practical rationality. It follows that human feeling is not only the mode of awareness of ourselves as practical agents—the constitution and making manifest of the self as a rationally desiring being, i.e., embodied, free, and efficacious—but equally the mode of awareness of other persons, as well as to our shared forms of activity. Thereby, moral respect, as the mode of self-consciousness which grounds our awareness of ourselves and others as moral beings, is constitutive of practical agency as such.

We can further see how all particular characteristically human feelings are manifestations of this unique mode of *self*-consciousness. The capacity to sustain oneself as a practical agent, i.e., as a free, embodied and efficacious being, requires intervening in

[14] For the definitive articulation of this idea see Steven Engstrom, *The Form of Practical Knowledge, A Study of the Categorical Imperative*.

the world in particular ways. This in turn requires assessing the *moral* status and relevance of the features of our environment, We've seen that feeling is the mode of awareness which reveal how conditions outside us would promote or hinder us from freely determining ourselves to act. Therefore, there is a need for a *receptive* capacity, which reveals the demands of reason on us, as embodied in the world around us, *on particular occasions*. This receptive capacity must be grounded in a self-understanding of oneself and others as free and therefore subject to moral demands, i.e., in the awareness afforded by the feeling of moral respect. For example, if I recognize a fellow human being is wronged, and feel indignation, this is nothing but a reflection and revelation of the moral fact that the situation ought to be remedied and moreover that it is perhaps up to me to do so. This feeling would reflect and reveal my understanding of myself as a moral being that is concerned with what is the right thing to do, and as one who is responsible for her actions.

As for feelings that are not obviously moral, these too, are to be understood as revealed to us in relation to our moral being, i.e., *through* moral respect. To be aware of how conditions outside us would promote or hinder us from freely determining ourselves to act, we need likewise to be aware of, but not thereby be determined by, how conditions outside us might affect us not qua being embodied, not self-sufficient, vulnerable.

§10. CONCLUSION

Moral respect should be understood as the distinctive *human* capacity for feeling. This should not be understood as a standalone faculty of feeling, whose operations are intelligible on their own. Instead, feeling in a practical agent is to be understood as a way of being morally self-consciousness: the way in which we know ourselves as free, embodied and efficacious. This is, in turn, a mode of receptive, moral, awareness of a rational agent whereby he becomes aware of how her own activity as well as the activities of others stand with respect to her essential end: to act for ends she sets for herself and in accordance with principle. Thus, moral respect, the distinctive human capacity for feeling, emerges as the form of self-consciousness *constitutive of practical agency*. [15] With this, Kant's treatment of moral respect reveals that *our* form of sensibility, human feeling, not only does not oppose but is the practical *embodiment* of reason.

[15] Carla Bagnoli has similarly argued that respect is "the emotional attitude that is constitutive of rational agency" in "Emotions and the Categorical Authority of Moral Reasons" (2011), 33. Likewise, Oliver Thorndike, attending to the systematic role played by moral respect in Kant's account, has claimed that "moral feelings are *essential* to autonomous agency, –not merely epistemological or motivational means to moral ends. […] Moral feelings are dispositions that should be cultivated for their *own* sake (their cultivation is *morally* obligatory), because they are essential –not merely supportive– to autonomous agency. They are not optional instrumental means that facilitate moral action or help to ward off temptations to trespass the moral law." "Kant's Transition Project" presented at Third Biennial NAKS Meeting, Emory University, May 2016. I take my account to be a further determination of these insightful interpretive suggestions.

Belief, Intention, and Deliberation

Greg Antill

Abstract: One of the central challenges in the philosophy of action involves explaining the relationship between an intention to perform an action, and the belief that you will perform that action. In this paper, I survey the two general schools of thought about this relationship – the *cogntivist* position, on which we identify an agent's intention to act with the belief that she will act and the *inferrentialist* position, on which an agent infers the belief that she will act from her intention to act – and the competing pressures that have pushed philosophers toward each view. I then propose an alternative picture which can accommodate the competing pressures for both sides and also presents an independently attractive picture of the relationship between belief and intention. Rather than identifying the belief and the intention, I argue that we should instead identify the *reasoning* by which we arrive at each.

One of the central challenges in the philosophy of action involves explaining the relationship between an intention to perform an action and the belief that you will perform that action.[16] In this paper, I will survey the two general schools of thought about the relationship – the cognitivist and inferrentialist positions – and the competing pressures that have pushed philosophers toward each view. I will then propose an alternative picture, one which I think can accommodate the competing pressures for both sides and one which also presents an independently attractive picture of the relationship between belief and intention. The position I want to argue for is a partial hybrid – rather than identifying the belief and the intention we should instead identify the *reasoning* by which we arrive at each: in the appropriate circumstances, the same reasoning can give rise to two distinct attitudes, an intention and belief, with the same propositional contents.

My argument will proceed in three steps. I will begin by noting certain structural similarities among the pressures for and against identifying belief with intention. I will argue that one could resolve both sets of pressures if one could have two distinct mental states that issue from one and the same piece of reasoning. In the second part of my argument, I will address a potential problem with such a solution: that in requires a blurring of the conceptual distinctions between practical and theoretical deliberation. I will conclude by arguing that this new view manages to carve out a genuine middle ground between the inferrentialist and cognitivist view without collapsing into one or the other.

§I. Cognitivism and Inferrentialism

[16] While I discuss the relation between beliefs and future intentions, much will also apply to the relationship between present-tense beliefs and intentions-in-acting. For some of the nuances involved in the relationship between these two sets of issues, see Falvey (2000).

That one might think the two mental states are related has a variety of sources. One is the similarity of their representational contents, marked in our surface-level grammar. "I will go for a walk" can equally well be an expression of the belief that one will go for a walk, or of one's intention to go for a walk.[17] This surface-level grammar is mirrored by a deeper connection between my acting intentionally and my knowing, or at least believing, that I am acting. A marker (perhaps even a sufficient condition) of a certain action being un-intentional – say my humming aloud in the office – is that I did not realize I was doing it.[18] This connection between my action being intentional and my believing that I am acting seems to be a connection in need of explaining, and one explanation would be in terms of a relationship between my intention and belief.[19]

A second pressure has to do with the immediacy and authority of first-personal belief about our actions.[20] The sorts of considerations on the basis of which I believe that I will act appear to differ sharply from the sorts of considerations on the basis of which I believe others will act. Unlike my predictions about the actions of others, which is based on behavioral or psychological evidence, my beliefs about my own actions are non-observational. Following Anscombe, many have argued further that we seem enjoy a special sort of *practical knowledge* with respect to beliefs about or own actions: my beliefs about my own actions are "justified, if at all, by a reason for acting, as opposed to a reason for thinking them true."[21] Since these are the very same considerations for which I come to intend to act, the puzzling phenomenon of practical knowledge is also to be potentially explained by some important relationship between the intention and the belief.[22]

There are two general schools of thought as to how best to explain these connections. The first position, motivated by the sorts of pressures sketched above, explains the connection by positing a very close relationship between intentions and predictions, namely the relationship of identity. This is the "cognitive" view of intention, which holds, at least in its simplest forms, that your intention to φ just is your belief that you will φ.[23] The explanation for why you must believe that you are acting, when acting intentionally, and why the considerations for which you predict are so similar to the reasons for which you intend is that your intention and your belief that you are or will act are one and the same.

Many have remained unpersuaded. This is in large part the result of equally strong pressures in the opposite direction to see intention as distinct from belief. One strong set

[17] Anscombe, (1957): §2
[18] Anscombe, (1957): §6. That belief, under some intentional description, is a *necessary* condition for intention is questioned in Davidson (1971): 50.
[19] This is a point made much of in the cognitivist account of Setiya, in a series of works: Setiya (2003); Setiya (2007); Setiya (2008); Setiya (2011).
[20] A point whose importance is highlighted by Anscombe (1957) and Hampshire (1959). For more recent development of this line of thought, see Velleman (1989); Moran (2001); and Wilson (2000).
[21] Anscombe (1957): 6. Anscombe is in turn following Thomas Aquinas' (ST IaIIae q3, a.5) conception of practical knowledge as "the cause of that which it understands."
[22] A point made much of in the cognitivist account of David Velleman. See especially Velleman, (1989) Ch. 3
[23] Perhaps the most ambitious attempt to give such an account is that of Velleman (1989). Others support some weaker version, on which intention is partly constituted by belief, (see e.g. Harman (1979) and Setiya (2008))

of pressures to keep intentions and beliefs about our actions distinct stems from the different roles the two types of mental states play in our larger mental architecture. A belief that I will φ may affect my inferences and actions in a very different way than an intention to φ.[24] My belief that I will smoke might prompt me to avoid convenience stores and so avoid inevitable temptation; my intention to smoke, in contrast, will prompt me to go to the convenience store as a means of satisfying the temptation to which I have already given in. My intention to go to the gym might be formed precisely *because* of my belief that I will likely succumb to laziness, in hopes of bolstering my resolve.[25]

A second and important set of pressures to keep our intention and belief distinct has to do with the evaluative conditions for believing and intending. My belief that I will act is successful in case it is true; my intention to act is successful in case it was choiceworthy or good.[26] My intention to act might thus count as correct even if my action is unsuccessful, provided I have good reason for acting, though my belief that I will act is incorrect. Conversely, I might, in acting immorally, succeed in believing correctly, but fail in intending well. Insofar as these standards of correctness for one attitude are exclusive, it seems as though the difference in the potential evaluation of my intention and belief provides further pressure to keep the two attitudes distinct.[27]

Thus the second, inferrentialist school of thought, on which the relationship between intentions and beliefs is a much weaker one. On the inferrentialist view, an agent's intention to φ is fully distinct from her belief that she will φ. What accounts for the connection between belief and intention is not some fact about the nature of the two mental states, but an epistemic connection: the intention can provide evidence for, and thus the basis to infer a belief about, the fact that one will φ.[28]

This view satisfies our intuitions about the different roles of an intention and belief that we will act, while going at least some way in explaining the connections between intention and our beliefs which had made cognitivism attractive. We will usually have a belief that we will act when we intend to act, because we will usually be in a position to easily infer from our intending to act that we will act. And we can account for some of the 'non-observational character' of our beliefs about our actions. For while our beliefs are predicted from evidence, like any other, we do have special first-personal access to the intention, on the basis of which we infer the belief that we will act. The apparent immediacy and authority of our first-personal beliefs about our actions can be off-loaded onto the the special authority and immediacy of our self-knowledge of our own minds.

But the inferrentialist view captures the epistemic phenomenon imperfectly. First, it fails to fully capture the necessary connection which obtains (at least in certain central cases) between belief and intention. If an agent infers that she will act from her

[24] Bratman (1987)
[25] Holton (1999); Holton (2009)
[26] A difference sometimes cached out in terms of differences in direction of fit (see, e.g. Humberstone (2002); Anscombe (1957)) and differences in terms of constitutive aim or function (see e.g. Velleman, (2000); Wedgewood (2002); Hieronymi (2005); Shah (2003); Burge (1999))
[27] Though less prominent in the current literature, this second response is in fact the oldest objection to cognitivism, going back to Aristotle NE 3.2
[28] Grice (1979); Paul (2009).

intention to act, there is always the possibility of her failing to draw the inference, however straightforward it might be. But, in at least many, if not all, cases of intentional action, it appears as though there is no space for me to fail to believe that I so act. The lack of space suggests that there are not two separate inferences – one practical inference to what I ought to do and another, from my intention to act to what I will do – but one.

Second, the inference fails to face squarely with the special access and authority an agent has in beliefs about her actions. Though I may be in a better position to know my intentions than you are, you could just as easily infer from the fact that I intend to φ to the fact that I will φ, as I can.

Most implausibly, on the inferrentialist view, distinctive knowledge of one's actions would be impossible for non-reflective agents, like animals and children. Since such agents lack the concepts necessary to represent intentions, they will not be able to infer their actions from the fact that they intend to act. But such agents *do* seem capable of distinctively first-personal knowledge of what they are doing, just like more sophisticated agents.[29] Though children and animals do not infer that they will act from their intentions, they do not appear to need to learn about their actions observationally, as a third-person observer might.

§II. Motivations for a New View

These objections are likely not conclusive – defenders of Inferentialism and Cognitivism would have much to say in response. I bring them up here for two reasons. First, I bring up these objections to motivate the appeal of some third, less problematic view – insofar as we can articulate a solution that avoids all these complications, this would be strong reason to adopt such a solution. But second, and more importantly, I also bring up these objections in order to look at the *structure* of the kinds of problems Inferrentialism and Cognitivism face, which I think reveals the shape that such a solution might take.

Inferrentialism and Cognitivism result from these competing pressures which appear to push two ways. Identifying intentions with action-beliefs will result in a prima facie violation of our intuitions about the different roles the two types of mental states play in guiding and regulating our activity, while keeping them distinct involves a prima facie violation of our intuitions about the close link, within a first personal point of view, between intending and coming to believe that you will act. As with many such situations, each view draws its strength primarily from the perceived weaknesses of the alternative.

Notice certain similarities in the form of these pressures for and against identifying belief with intention. The first group of pressures which push toward identifying belief with intention are all what we might call "upstream" pressures – pressures associated with the effects of a subject's prior mental states and processes on a propositional attitude. They stem from the apparent connections between coming to intend to act, and coming to believe that you will act. In contrast, the second group of pressures against cognitivism, and toward inferrentialism, are all "downstream" pressures – pressures associated with the effects of a propositional attitude on the subject's resulting thought and behavior. They are pressures that stem from the different causal,

[29] For further discussion, see O'Brien (2007).

explanatory, and evaluative roles that an intention and belief with the same relevant content might play in our mental architecture.

Notice too that you will resolve both pressures if you could have two distinct mental states that issue from one and the same piece of reasoning. The fact that they issue from the same piece of reasoning would explain why the belief and the intention share so many 'upstream' features – they are the result of the very same process. But the fact that they are still two distinct attitudes both preserves our intuitions about the distinction between belief and intention and explains their potentially divergent effects on our downstream mental lives. If it were possible to have two distinct metal states that issue from one and the same piece of reasoning, that would be the kind of solution with exactly the right structure to accommodate both sets of conflicting upstream and downstream pressures.[30]

The problem with this solution is that it appears to require a rejection of the close conceptual connections between attitude-types and the kinds of reasoning from which they normally issue. Part of what it is to be a belief that p, as opposed to some other attitude like an imagining that p or a desire for p, is that it be the kind of state which issues from or is partly constituted by the conclusion to a piece of *theoretical* reasoning about whether p is so. Whereas part of what it is to be an intention to φ, as opposed to some other attitude, is that it be the kind of state which issues from or is partly constituted by the conclusion to a piece of practical reasoning about what to do.

While not all beliefs and intentions need issue from some explicit piece of reasoning, the fact that they are the kind of attitude which issued from *some* process which functioned to track the truth, or determine what to do, respectively, is part of what explains why they play the distinctive functional role in our mental architecture that they do and why they are answerable to the distinctive standards of correctness to which they are.

A theory on which an intention and belief both issue from the same piece of reasoning will thus face the following dilemma: for both a belief and an intention to issue from the same piece of reasoning, it looks like it would have to be either a piece of theoretical reasoning aimed at answering some theoretical question of what is so, or a piece of practical reasoning aimed at some practical question of what to do. If it is the former, you have not successfully come to an intention; if it is the latter, you have not successfully come to a belief.[31]

How could this dilemma be avoided? If states are typed by the kind of reasoning from which they issue, and kinds of reasoning are typed by the kind of inquiries at which they are directed, then we would need some piece of reasoning, the resolution of which amounts to the settling, at the same time, of both a practical question of what to do and a theoretical question of what I will do.

I want to suggest that, with respect to beliefs about our own actions, we can do just that. When we have control over the state of affairs we are deliberating about, we can answer a theoretical question of whether some state of affairs is so by answering a

[30] In this paper I will understand reasoning broadly, including not just explicit conscious train of thought, but any psychological transition by which we come to form, revise, or sustain an attitude for reasons.
[31] A problem familiar from the wrong kind of reasons literature, emphasized by Hieronymi (2005).

practical question of whether it should be so.[32] In doing so, I will argue, we are not doing anything remarkable. That it might seem otherwise is the result of impoverished thinking about the nature of theoretical reasoning. I will show that when theoretical reasoning is properly understood, we can see that believing that we will act on the basis of reasons for acting is, in all important respects, structurally identical to more familiar cases of believing for reasons on the basis of evidence.

§III. Decision as Theoretical Inference

It should be uncontroversial that when we inquire into whether some state of affairs is so, we often do so by means of employing some mixture of inference rules and premises. But I think the richness and variety of ways we engage in theoretical reasoning is often obscured. Just as it is a mistake to think of practical reasoning as the mere adding up of reasons or desires with a certain valence and magnitude, it's also a mistake to think of all theoretical reasoning as the simple tallying up of various pieces of evidence, each of which shows the conclusion more or less likely to a stronger or weaker degree.

Theoretical reasoning is not such a monolithic phenomenon. In fact, there is a great variety of kinds of considerations which we might employ in theoretical reasoning, and a great variety of ways in which we might bring them to bear on the question of whether some proposition is so. To give just a few examples:

(i) Sometimes we answer theoretical questions by means of induction from particular cases. We might infer from green emerald facts to the belief that all emeralds are green.[33]

(ii) Sometimes we answer theoretical questions by means of abduction from explananda. We might infer from dinosaur track facts to the belief that dinosaurs existed.

(iii) Sometimes we answer theoretical questions by predicting from causal antecedents. We might infer from storm cloud facts to the belief that it will rain.

(iv) Sometimes we answer theoretical questions by constitution from grounds. We might infer from the occurrence of certain troop movement facts to the belief that Germany is at war with France.

[32] This way of proceeding is deeply indebted to the work of Moran (2001) on self-knowledge. I see this account as broadly congenial with, and as providing a possible explication of, his theory of self-knowledge (though not, I think, one Moran himself would accept). Just as we can control what we do by deciding to do it, we can control what we believe, by making up our minds about what is so.

[33] Here I assume that enumerative induction is a distinct form of inference from abduction, contra Harman (1965). The essential thing for the present discussion, however, is not *which* of these methods are the truly distinct kinds of inference, but that there *are* distinct kinds of inference, operating on distinct kinds of grounds.

The list is not exhaustive. There are lots of distinct kinds of inferences, operating on lots of distinct kinds of grounds. Each is a way of answering a theoretical question about whether p is so, by way of pursuing some other question of evidence, explanation, constitution, or causation.

When you have (or take yourself to have) control over what happens, I want to suggest that there is a further available method for determining whether some state of affairs is so:

(v) Sometimes we answer theoretical questions by deciding what will happen, on the basis of our practical reasons. You might infer from the fact that she killed your father to the fact that you will get revenge.

The first four methods are all ways of committing oneself to the truth of p (and so coming to believe that p) by *discovering* whether p is so. What thinking about the case of belief about action reveals is there is another way to commit oneself to p's being so, by *deciding* that it will be so.

In such circumstances, we will have answered, at once, both a practical question of what to do and a theoretical question of what we will do. We will thereby have concluded a deliberation which was both practical and theoretical, and so have come to form both an intention to act, and a prediction that we will so act, each with differing standard of correctness, and potential divergent effects on our downstream mental lives.

That this is possible requires no re-thinking of the nature of belief or theoretical reason. Figuring out that you will act a certain way by deciding that you will act a certain way is just one more method among our varied repertoire for engaging in theoretical deliberation about what is so. Of course, (v) is distinctive from (i)-(iv) in the following respect: the first four are all ways of answering the question whether p by discovering whether p is so. I am treating p as some independent state of affairs, outside my control, where I am to determine whether it obtains by addressing myself to some question which bears on whether p is likely or probable. In contrast, in (v) I am answering the question of whether p by bringing p about.

But I can see no reason why such reasoning ought not count as genuinely theoretical. Each of these methods seem, in the right circumstances, like equally reasonable ways of determining whether some state of affairs is so. A proposition is no less true because I decided to make it so than because I discovered it was so. Moreover, if we really are in control with respect to the state of affairs we are forming a belief about, it will be *good* theoretical reasoning. If we conclude that some state of affairs will obtain, by deciding that state of affairs will obtain, we will be right.

If it is a reasoning which can be directed toward answering a question of what will happen, and it is a reliable means of determining what will happen, I see no reason why this unique causal direction should matter. Deciding to φ is, in all relevant respects, just one more means of answering the question of whether you will φ, along with all the other evidential means we have at our disposal.

§III. Conclusion

I will conclude by arguing that this view manages to carve out a genuine middle ground between the cognitivist and inferrentialist view, without collapsing into one or the other. In doing so, it is able to avoid the problems facing each of these views.

Recall that a successful treatment of the connection between belief and intention must walk a careful line. One the one hand, a successful treatment must be able to explain why – in cases of difficult action or foreseen weakness of will – we can intend to act but fail to believe that we will, and vice-versa. At the same time, it must explain why, in many central cases of action, we necessarily believe that we are acting as we intend.

The single-reasoning view can explain, as a theory on which we infer from our intentions to the belief that we will act cannot, why a belief that I am acting is sometimes necessary condition for intending that I act. If the decision to act is the inference by which you conclude that you will act, there is no room for failing to draw the appropriate epistemic connections between intending and acting. The conclusion of your deciding whether to φ just is your conclusion that you will φ. And it can explain, as a theory on which we infer the belief that we will act from an intention to act cannot, our distinctive first-person access. Because I am not in a position to control how another agent acts, I cannot come to believe that she will act by deciding how she should act.

But the single-reasoning view can *also* explain, as a view which identifies belief with intention cannot, why, in certain special cases, like cases of difficult action or weakness of will, we can believe that we will act without an accompanying intention to act, or intend to act while doubting that we shall succeed. Since we can come to believe in many different ways, the single-reasoning view allows that we *can* form a belief about what we will do by predicting, and not deciding. Indeed, we have the capacity to settle the question of whether a proposition is so by deciding that it is so, only when we take ourselves to have control over what will happen. And in cases of difficult action or weakness of will, we may take ourselves to lack such control, and so the question of what we will do and the question of what to do can come apart. In such circumstances, we might decide to act, but predict that we will not, and so arrive at an intention to act, absent an accompanying belief, or vice-versa.

Bibliography

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.

Aristotle, (1985). *Nicomachean Ethics*. (*Trans.* Irwin, T.) Indianapolis, Ind: Hackett Pub. Co.

Aquinas, T. (2008). *Summa Theologiae* (2 ed.). NewAdvent.org

Baier, A. (1970). 'Act and Intent,' *Journal of Philosophy*, 67: 648–658.

Boyle, Matthew (2011). 'Making up Your Mind' and the Activity of Reason. *Philosophers' Imprint* 11 (17).

Bratman, M. (1987). *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.

Davidson, D. (1971) 'Agency,' reprinted in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, pp. 43–61.

Evans, Gareth (1982). *Varieties of Reference*. Oxford University Press.

Falvey, Kevin (2000). Knowledge in intention. *Philosophical Studies* 99 (1):21-44.

Grice, H. P., (1971). 'Intention and Uncertainty,' *Proceedings of the British Academy*, 5: 263–279.

Hampshire, S. (1959). *Thought and Action*, Notre Dame, IN: University of Notre Dame Press.

Hieronymi, Pamela (2005). The Wrong Kind of Reason. Journal of Philosophy 102 (9):437 - 457.

-- (2009). Two kinds of agency. In Lucy O'Brien & Matthew Soteriou (eds.), Mental Action. Oxford University Press. 138–162.

Holton, R. (1999). "Intention and Weakness of Will," *Journal of Philosophy* 96 (5):241-262.

-- 2009, *Willing, Wanting, Waiting*, Oxford: Oxford University Press.

Humberstone, I. L. (1992). Direction of fit. *Mind* 101 (401):59-83.

Marusic, Berislav (2012). Belief and Difficult Action. *Philosopher's Index* 12 (18):1-30.

Moran, Richard A. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.

Paul, S. (2009). 'How We Know What We're Doing,' *Philosophers' Imprint*, 9: 1–24.

Setiya, Kieran (2011). "Knowledge of Intention." In Anton Ford, Jennifer Hornsby & Frederick Stoutland (eds.), *Essays on Anscombe's Intention*. Harvard University Press. 170 - 197.

-- (2007). *Reasons without Rationalism*, Princeton, NJ: Princeton University Press.

-- (2008). "Practical Knowledge." *Ethics* 118 (3):388-409.

-- (2003). "Explaining Action." *Philosophical Review* 112 (3):339-393.

Shah, Nishi (2003). How truth governs belief. *Philosophical Review* 112 (4):447-482.

Thompson, M., 2008, *Life and Action*, Cambridge, MA: Harvard University Press.

Velleman, David (1989). *Practical Reflection*. Princeton University Press.

-- (2000). On the aim of belief. In , *The Possibility of Practical Reason*. Oxford University Press. 244--81.

Wedgwood, Ralph (2002). The aim of belief. *Philosophical Perspectives* 16 (s16):267-97.

Wilson, George (2000). "Proximal Practical Foresight." *Philosophical Studies* 99 (1):3-19.

Contingency and Integrity

Joshua Tignor

Abstract: common concern for contemporary moral theories has to do with the maintenance of agential integrity. In order to avoid making moral agents overly intellectual, moral theories often resort to separating the part of us that engages in moral philosophizing from the part of us that makes everyday decisions. The worry with this move is that by dividing the agent into two distinct selves that have two distinct takes on the world, one might threaten the internal coherence or integrity of the agent. I want to take this particular concern with agential integrity and look at a somewhat recently proposed form of constructivism called Humean Metaethical Constructivism. I will focus on Sharon Street's account of how moral agents can come to terms with the contingency of their moral commitments without undermining the unique normative authority associated with those commitments. I will argue that Street's response to this worry seems to threaten the internal integrity of the moral agent. From here I will present a possible response on behalf of the Humean that suggests a kind of internal alienation that, if possible, doesn't seem too threatening to the internal integrity of the agent.

A common concern for contemporary moral theories has to do with the maintenance of agential integrity. In order to avoid making moral agents overly intellectual, moral theories often resort to separating the part of the agent that engages in moral philosophizing from the part that makes everyday decisions. By doing this, proponents of contemporary moral theories avoid making everyday moral agents into moral philosophers. The worry with this move is that by dividing the agent into two distinct selves that have two distinct takes on the world, the internal coherence or integrity of the agent might be threatened.

One way of thinking about this move is in terms of a separation of self.[34] To avoid over intellectualizing moral agents, philosophers separate the philosophical self from the practical self.[35] Another way of thinking about the worry is in terms of perspectives. The practical perspective takes itself to be rendering judgments about the world while the philosophical perspective takes itself to be considering the nature and justification of the principles that ground the judgments issued from the practical perspective. The worry is that it now seems that moral philosophers have opened

[34] A good representation of this is Hare's two level utilitarianism in *Moral Thinking* (1981).

[35] You might also think of the issue here in terms of alienation. That is, in order to avoid the over-intellectualization of moral agents, contemporary moral theories require agents to alienate themselves from the philosophical perspective when acting from the practical perspective, and to alienate themselves from the motivating features of their practical self when occupying the philosophical perspective.

themselves to the possibility of these perspectives having conflicting or irreconcilable takes on the world, and thus the possibility of a *severe* form of internal alienation.[36]

Now this worry has not gone unnoticed. Many scholars have attempted to defend their preferred moral theory from such worries.[37] That said, the present paper is not an attempt to contribute to this discussion concerning first-order normative theories. Instead, I want to take this particular concern with agential integrity and look at a somewhat recently proposed form of constructivism called Humean Metaethical Constructivism.[38] Humean metaethical constructivism argues that the substantive content of an agent's moral commitments is justified according to other, *contingently* given substantive value commitments. Under this account, it could have been the case that, given different substantive value judgments upon entering agency, the substantive content of an individual's moral judgments would have also been different.

At this point you might think that making the justification of an agent's moral commitments contingent might lead to an undermining of the uniquely categorical nature of the moral reasons those commitments generate. After all, if my moral commitments and their justifications could have been otherwise (i.e. are contingent), then why do I let them command me despite what I want? This is important because the particular conception of Humean metaethical constructivism under consideration here claims that the agent can come to terms with the contingency of their moral commitments.

The focus of the present paper is on Sharon Street's account of how Humean moral agents can acknowledge this contingency without having it undermine the unique normative authority of their commitments. I will argue that Street's response to this worry seems to threaten the internal integrity of the moral agent. The Humean agent seems to be engaging in some form of internal alienation if they are capable of both acknowledging the contingency of their moral commitments and also viewing those commitments as, in a sense, commanding categorically. From here I will present a potential response on behalf of the Humean that suggests a kind of internal alienation that, if possible, doesn't seem too threatening to an agent's integrity.

I. Love, Morality, and Contingency

Humean metaethical constructivism holds that our moral commitments often command us apart from what we most want and that the justification for an agent's moral commitment that makes this command is a contingent matter. The Humean moral agent acknowledges that they could have come into agency with a different set of substantive value commitments and thus developed a different set of moral commitments. They can imagine a world where, things having happened a bit differently, they were a utilitarian instead of a Kantian, or one where they had no moral commitments at all.[39]

---

[36] Think of the paradox of hedonism. As a moral theory, it seems like the only way a hedonist can be a good hedonists is by not being a hedonist in some situations. Friendship is one example. For a friendship to bring me pleasure, I cannot be motivated by a concern for my pleasure.

[37] Railton (1984); Hare (1981); Williams (1981; 1988); Herman (1983); Swanton (1997); Martinez (2011); Keller (2011).

[38] Street (2012); Lenman (2010).

[39] Street (2012), pp. 53.

Street claims that our relationship with moral commitments, and their contingency, is analogous to the contingency associated with long-term, committed relationships. She says that unless one is a hopeless romantic, it will not be difficult for an individual to acknowledge the contingency of their commitment to the person they consider to be their life-long love. In fact, that individual can actually imagine a world where things went a little differently at some point in their life and they did not meet the person they are currently with, and instead, met, fell in love with, and married someone different. There relationship is contingent insofar as it could have been otherwise and there is no particular justifying reason for why they ended up with their current partner instead of someone else.

Importantly for Street's argument, acknowledging this contingency does nothing to undermine the love and devotion the individual has for their current partner. In fact, Street suggests that the contingency of the relationship might actually make it dearer. Of all the possible worlds where they might have ended up with someone else, they ended up in this one with their current partner. What is important to take from this example is that the contingency of the relationship does not seem to undermine the love and devotion characteristic of such a relationship.

Consider Sam and Taylor. Suppose these two individuals have the kind of long-term, committed relationship described above. The love and devotion constitutive of this relationship manifests itself in the form of normative commitments. There are certain things Sam and Taylor ought to do given the relationship they have agreed to have with each other. And importantly, these normative commitments often issue commands opposing what they most want as individuals. Despite the fact that what Taylor most wants is to see a Broadway show this weekend, he knows that he should go with Sam to his family reunion. And importantly, Taylor acknowledges that his meeting Sam and forming such a relationship was to a large extent dependent upon factors outside of his control. It could have been the case that he didn't meet Sam given some difference in these factors, and yet, Taylor still loves Sam nonetheless. He can acknowledge that he could have met and fallen in love with someone else and that there is no justificatory reason apart from certain contingencies that explains why he is with Sam instead of someone else. And despite all of this, he still holds that there are certain things he ought to do given his relationship with Sam.

Street's suggestion is that we might understand our relationship with morality in a similar way. The Humean moral agent knows that they could have developed a different set of moral commitments. They can imagine a world where, things having happened a bit differently, they have drastically different moral commitments, or none at all. Importantly, and like the long-term, committed relationship, acknowledging this contingency does nothing to undermine the unique normative authority associated with their current moral commitments.

II. Contingency and a Worry

Trying to reconcile contingency with a unique normative authority could be understood as a potential source of anxiety for the Humean. It seems plausible that acknowledging that the justification for her moral commitment *could have been*

otherwise and that there *is no particular reason why it is justified this way as opposed to some other* could undermine its capacity to command the agent apart from what she most wants. If the justification for her moral commitment is a contingent matter in this sense, then why should she grant it greater normative authority then her other substantive value commitments?

Consider for a moment Susan. Let's say that what Susan most wants at this moment is a chocolate milkshake, and importantly, Susan doesn't have enough money buy one. Also let's assume that Susan is a Kantian and is committed to respecting the autonomy of others. Regardless of the fact that what Susan most wants is a chocolate milkshake, Susan is not going to beat up her office mate and take his money to buy one at the Student Union Building. Her moral commitment to respecting the autonomy of others stops her from acting on what she most wants.

But now, as a Humean moral agent, Susan finds herself in a peculiar state of mind. Susan can ask herself why it is she should not beat up her office mate and steal his money. Given her Kantian commitment, Susan would claim that it is because doing so would not respect the autonomy of her office mate. She can continue her line of questioning and inquire as to why she should respect the autonomy of others. At this point Susan is searching for what justifies the normative authority of her current moral commitment. And it is at this point that our Humean moral agent encounters a peculiarity. Susan now sees that the justification for her commitment to respecting autonomy depends upon other, contingently given value commitments, and thus could have been otherwise. She knows that her moral commitment could have been otherwise and that there is no principled reason why it is this way instead of some other. So, if it could have been otherwise, why should she see herself as bound to upholding her moral commitment as opposed to doing what she most wants?

Now this would not be much of a worry if Humean constructivism did not support the claim that moral commitments command us apart from what we most want. If moral commitments did not have this unique normative authority, then the contingency of their justification undermines nothing of importance. Consider what Street says when discussing the above worry,

> [On] the view I'm suggesting, it's constitutive of being a moral agent that one take certain requirements (of a certain characteristic content concerning the equal treatment of others) to be binding even if carrying them out goes against certain large aspects of one's evaluative nature—including what feels easiest, what is pleasant, fun, what one finds most naturally appealing, and so on.[40]

Humean constructivism is committed to an understanding of moral commitments as having the capacity to command us apart from what we most want. Coming to terms with the contingency of our moral commitments still seems to present a potential problem for the unique normative authority associated with our moral commitments. Exactly how this unique normative authority is maintained in the face of contingency has to do with

[40] Street (2012), pp. 52.

the essential role moral commitments play in the formation of an agent's practical identity.

III. Normative Authority and Practical Identity

Arguing along the same lines as Korsgaard (1996)[41], Street claims that moral commitments are an essential aspect of a moral agent's practical identity.[42] Practical identity, in the sense intended here, can be understood as a description of one's self that gives rise to reasons for action for the person whose identity it is. Being a member of the Roman Catholic Church, an atheist, a woman, transgender, or in a committed relationship are all practical identities that generate obligations for individuals.

Street is suggesting that moral commitments are fundamental to any moral agent's practical identities. That is, moral commitments constitute another practical identity upon which all of our other contingent practical identities are built. These commitments define who we are in our own eyes. Losing such commitments entirely would be to vanish or perish in a deeply intuitive sense. If a particular agent were to lose their moral commitments, that moral agent, as defined by those commitments, would no longer exists.

Street argues that it is because of the fundamental role moral commitments play in a moral agent's practical identity that they maintain their capacity to command apart from what is most wanted. They are so fundamental that the agent cannot look at those commitments and at the same time think that they are a contingent matter. The fact that it could have been otherwise and that there is no particular reason why they are the way they are is, in a sense, inaccessible from the first-person, agential perspective. Contingency doesn't cause problems for the Humean moral agent because they simply cannot acknowledge the normative authority of their moral commitments while at the same time thinking that those commitments and their justification are a contingent matter.

IV. Contingency and Practical Identity

With this account of the relationship between moral commitments and an agent's practical identity in mind, I want to look back to Street's analogy between morality and long-term, committed relationship. Recall that, for Street, an individual in a long-term, committed relationship can acknowledge the contingency of that relationship. They can actually imagine a world in which they ended up with a different partner, and importantly, coming to terms with this contingency doesn't undermine their commitment to their current partner. I want to suggests that what allows the individual to come to terms with the contingency of this relationship has to do with fact that in each of those possible worlds where things went differently, they are assuming the same practical identity.

Consider again Sam and Taylor. Taylor can actually imagine a world in which he ends up in a long-term, committed relationship with someone other than Sam. The

[41] Korsgaard (1996), pp. 101-102.
[42] Street (2012), pp. 53.

reason Taylor can do this is because the individual in the other world is still him in terms of practical identity. He is imaging himself, as constituted by his current moral and other substantive value commitments, in another world having met someone different. Another way of putting it is that Taylor can *practically identify* with the individual in the other world. This is what allows him to come to terms with the contingency of that relationship. In each case where it turned out differently, it turned out differently *for him*.

My suggestion is that a moral agent under the Humean picture cannot intelligibly conceive of a possible world in which they have different moral commitments, and thus cannot come to terms with the contingency of those moral commitments. Consider again Susan. The justification for the moral commitments that constitute Susan's practical identity are, according to the Humean picture, contingent upon other substantive value commitments. That said, according to the picture of practical identity presented earlier, what allows such commitments to command apart from what Susan most wants is the fact that these commitments are fundamental to Susan's practical identity. They are so fundamental that Susan cannot look at them and at the same time think that there presence is a contingent matter. Their unique normative authority is generated by their fundamentality.

It is this last claim that seems to cause problems for the Humean moral agent and their attempt to come to terms with the contingency of their moral commitments. It seems that this fundamental relationship between an agent's moral commitments and their practical identity might actually *prevent* them from coming to terms with the contingency of their commitments.

Recall that what allowed Taylor to acknowledge the contingency of his committed relationship was the fact that he could intelligibly conceive of a possible world where he met, fell in love, and married an individual other than Sam. Furthermore, it seems that based on the account of practical identity championed by Street, forming such a conception, if it is to be meaningful to the relevant agent, requires that the individual be able to practically identify with the individual in the other possible worlds. That is, for Taylor to come to terms with the contingency of his relationship, it must be the case that the Taylor in the other possible world is, in some fundamental sense, the same as the Taylor in the actual world. Taylor can conceive of a world where something goes differently for Taylor if the person in the other world *is* Taylor is some deep fundamental sense.

Unfortunately, this cannot be the case for our Humean moral agent Susan. Given the relationship between moral commitments and practical identity discussed above, a change or loss of moral commitments results in a vanishing of the fundamental practical identity that defined original moral agent. Coming to terms with the contingency of her moral commitments, if we are to stay true to the analogy, requires that Susan be capable of conceiving of a possible world where she has different moral commitments. But this is impossible for Susan. Conceiving of such a world would mean letting go of her current moral commitments and adopting another set. But once this change in moral commitments has occurred, she is no longer concerned, in terms of practical identity, with herself. Another way of putting it is that Susan cannot conceive of a possible world where Susan has different moral commitments because *Susan* only exists when constituted by her current commitments.

Coming to terms with the contingency of one's moral commitments means being able to conceive of a possible world where one has different commitments. Conceiving of such a world *seems* impossible for the Humean moral agent given the fact that letting go of one's moral commitments results in that agent perishing in terms of practical identity. I stress "seems" here because in the next section I suggest a potential response on behalf of the Humean position.

V. Contingency and Alienation

Consider for a moment that the Humean constructivist's project is the most plausible account on hand. It now seems as if Susan can say that her moral commitments and their justification are both contingent and non-contingent. As an active moral agent, Susan now seems to be in a strange position. I want to suggest that this strange position can be accounted for in a seemingly unproblematic manner.

One way to approach the issue is in terms of perspectives. We might say that when acting as moral agent, from a practical perspective, Susan cannot view her moral commitments and their justification as being contingent. But, in a cool moment, Susan can step back from her practical perspective and acknowledge the contingency of her moral commitments from a critical or philosophical perspective. The Humean might suggest that Susan has access to, and can switch between, two different perspectives on the same phenomenon. Her philosophical perspective allows her to *look at and assess* her moral and other substantive value commitments that constitute her practical identities. From here she can acknowledge the contingency of her moral commitments and their justifications. She also has a practical perspective that is *constructed by* her moral and other substantive value commitments. From this perspective, Susan cannot view her moral commitments as justified contingently because of the fundamental role they serve to her practical identities.[43]

We can think of what's happening here in terms of *alienation*. It seems like the Humean moral agent, in order to acknowledge her moral commitments as justified when they command unique normative authority, must alienate herself from the fact that the justification of those commitments is itself contingent. Susan must, in a sense, close herself off from the fact that her moral commitments are contingent. The worry with this move is that it seems to pose a threat to the internal coherence or integrity of the Humean agent.

With that said, there might be reason to believe that not all instances of alienation are problematic. Consider a case used by Julia Markovits (2014).[44] In a war that is justified on certain humanitarian grounds, in order for the soldiers to be effective in

[43] This distinction between two perspectives is similar to the idea behind Hare's (1981) two-level utilitarianism. The only difference is that I am here referring to the "critical level" as the "philosophical perspective" and the "intuitive level" as the "practical perspective". The reason I'm choosing talk in terms of perspectives is because I think it makes more clear the kind of issue the Humean moral agent is dealing with when attempting to acknowledge the contingency of their moral commitments. It has to do with *taking* a certain moral commitment as being contingent from one perspective while at the same time *taking* that moral commitment to be non-contingent (categorical) from another perspective.

[44] Markovits (2014), Pg. 47.

achieving their goal, they ought not be motivated by a concern for humanity in general. The reason is because a concern for all humanity would make it hard, if not impossible, for the soldier to view their enemy *as an enemy*. One could say that the soldier *must* alienate herself from her concern for humanity in general in order to be effective. What is important here is that, at face value, there does not seem to be any glaring problems with allowing this kind of internal alienation.

The question now is whether or not the Humean moral agent is undergoing a similar kind of seemingly unproblematic alienation. Answering this question will involve acknowledging a subtle difference between the two cases. In the case of the soldier, she is alienating herself from the actual justification of her actions. For the sake of humanity, she must not be immediately concerned with the common humanity possessed by all individuals. In the case of the Humean moral agent, what we get is an individual alienating herself from *the fact* that the justification for her moral commitments is contingent. This is different from the soldier case because the soldier was alienating herself from the actual justification of her actions whereas the Humean moral agent is not alienating herself from what justifies her moral commitments, but from a fact about the way they are justified. This appeal to alienation does not seem obviously problematic, but, as I will suggest in the conclusion, it might not be that helpful.

VI. Conclusion

I have argued that a moral agent under Street's version of Humean constructivism cannot come to terms with the contingency of her moral commitments. This is because of the fundamental role Humean constructivism attributes to moral commitments in the constitution of an agent's practical identity. Losing such commitments would cause the agent to vanish in terms of their practical identity. Because they are so fundamental, the agent cannot look at them and at the same time think that their moral agency is a contingent matter. The problem is that, because of this, the Humean moral agent cannot imagine a possible world where they have different moral commitments. To intelligibly conceive of such a world would be to drop one's current set of moral commitments and adopt a different set. But, losing one's moral commitments means to vanish in terms of practical identity. The Humean moral agent cannot come to terms with the contingency of their moral commitments because doing so would require imagining their self with different moral commitments. And this is impossible according to the above account of practical identity appealed to by Street.

Now, it might be the case that there is an unproblematic instance of alienation occurring in the Humean moral agent. That is, in order for an agent to view her moral commitments being justified categorically, she must alienate herself from the fact that the justification for those commitments is a contingent matter. The soldier in the just war seems to make such alienation seem more plausible and unproblematic.

I want to briefly suggest a reason for thinking that this kind of alienation might not be problematic, but also might not be that helpful. The reason alienation might not be problematic is due to the fact that this particular fact about the justification of one's moral commitments does not seem to do any work for the agent in terms of motivation. It is simply a fact. It does not seem to give an agent reasons to act one way or another. This

being the case, alienating oneself from this fact does not seem to cause much trouble. If it doesn't do any work for the agent in terms of motivation, then why be concerned with maintaining epistemic access to it? That said, there is also a worry that such alienation might not be that helpful for the Humean. When switching between perspectives, it is arguably the case that the agent is not abandoning their moral commitments. Has the agent really put their self in a position to acknowledge the contingency of those commitments? Again, these are only tentative and underdeveloped avenues of inquiry on behalf of the Humean. Maintaining the plausibility of the Humean approach requires further and more in depth analysis of these questions.

References

Bagnoli, Carla. (2002), "Moral Constructivism: A Phenomenological Argument", *Topoi*, 21, pp. 125-138.

Hare, R. M. (1981), *Moral Thinking: Its Levels, Method, and Point*. OUP.

Herman, Barbara (1983), "Integrity and Impartiality". *The Monist,* 66(2), pp. 233-250.

Keller, Simon (2007). "Virtue ethics is self-effacing". *Australasian Journal of Philosophy*, 85(2), pp. 221–231.

Korsgaard, Christine M. (1996), *The Sources of Normativity*, Cambridge University Press

Lenman, James. (2010), "Humean Constructivism in Moral Theory", *Oxford Studies in Metaethics,* Vol. 5, Oxford University Press, pp. 175-193.

Markovits, Julia. (2014), *Moral Reason*, Oxford University Press.

Martinez, Joel (2011). "Is Virtue Ethics Self-Effacing?" *Australasian Journal of Philosophy*, 89(2), pp. 277-288.

Railton, Peter (1984). "Alienation, consequentialism, and the demands of morality". *Philosophy and Public Affairs*, 13(2), pp. 134-171.

Street, Sharon. (2010), "What is Constructivism in Ethics and Metaethics?", *Philosophy Compass,* 5:5, pp. 363-384.

Street, Sharon. (2012). Coming to terms with contingency: Humean constructivism about practical reason. *Constructivism in practical philosophy*, 40-59.

Swanton, Christine (1997). "Virtue Ethics and the Problem of Indirection: A Pluralistic Value-Centred Approach". *Utilitas*, 9(2), pp. 167.

Williams, Bernard (1981). "Persons, Character, and Morality". In James Rachels (ed.), *Moral Luck*. Cambridge University Press.

Williams, Bernard (1988). "The Structure of Hare's Theory. In, *Hare and Critics: Essays on Moral Thinking*. OUP.

Love, Reason, and the Highest Good

David Sussman

Virtue and happiness exist in an unstable relationship in Kant's thought. Kant spends much of his practical philosophy arguing that virtue is a matter of responsiveness to a moral law that is prior to and independent of the rational concern that we properly have for our own happiness. Yet Kant was never able to fully accept such a duality of practical reason. Throughout his mature works, Kant insists that there must be a necessary (if synthetic) relationship between virtue and happiness. Although virtue's importance is prior to any prudential concerns, Kant nevertheless holds that reason requires that virtue be "crowned" with happiness in a condition known as the Highest Good. For Kant, the Highest Good would be realized in a world where everyone, having attained perfect virtue, enjoys the happiness they thereby deserve just because they so deserve it. In such a state we would not only be fully virtuous and completely happy; it would also be the case that, had we been any less virtuous, we would have been proportionately less happy, as if by natural law.

Kant contends that since morality requires that we strive to inaugurate the Highest Good, we must have a kind of "rational faith" in the prerequisites for success in this project: i.e., the existence of God, the freedom of the will, and the immortality of the soul. Freedom and immortality are supposedly needed for us to morally perfect ourselves, in part because Kant holds that perfect virtue cannot be attained in any finite span of time. We must have faith in the existence of God as a power that can properly assess our virtue, and order the laws of nature so that we all receive are just deserts thereby. Kant offers a variety of different argument for the necessity of the Highest Good, all of which seem to be in tension with central commitments of his moral philosophy. In this 2 paper, I offer a defense of the Highest Good drawn from some remarks Kant makes in his late work, Religion within the Limits of Reason Alone. In the Religion, Kant claims that the Highest Good must be possible as the result of our efforts because human reason requires not just a law of action, but as an "object of love." I contend that the love that warrants rational faith should be understood as a special kind of inclination that Kant calls a "passion" (Leidenschaft). As creatures who must grow into their reason from a sensible starting point, human beings experience their lives in time as an endless series of passions, a series that, if it has the right form, will count as the attainment of pure practical reason. I argue that for this progress of the passions to count as the emergence of reason, it must be focused on and unified by a commitment to the Highest Good.

Kant offers his central argument for the Highest Good in the Critique of Practical Reason. In the second Critique, Kant argues that the validity of the law depends on the possibility of attaining the Highest Good. Here Kant argues that since morality is the condition of the value of any sort of happiness, we should understand virtue as the "worthiness to be happy". There is thus a necessary connection between the concepts of virtue and happiness, insofar as

> to need happiness, to be also worthy of it, and yet not to participate in it cannot be consistent it cannot be consistent with the perfect volition of a rational being that

> would at the same time have all power, even if we think of such a being only for the sake of the experiment. (5:110).

Kant then argues that any two distinct concepts that nevertheless stand in a necessary relation to each other count, logically, as an instance of the relationship of cause to effect. As Kant interprets this relation, either happiness must serve as the cause of virtue, in the sense that people become virtuous in the attempt to be happy, or virtue must be cause the of happiness, in that we come to enjoy the amount of happiness we do purely of the 3 degree of virtue we have attained. Kant rejects the first option, that we become virtuous for the sake of happiness, as being incompatible with the basic autonomy that he takes to be at the heart of morality. As a result, Kant concludes that virtue must be the cause of happiness, which it supposedly can only be if God exists to make the laws of nature bestow upon us our just deserts, and if there is a future life in which to enjoy them.

Kant claims that, in our intuitions about desert, we see the concepts of virtue and happiness being connected by necessity, and that two distinct concepts so connected logically count as a kind of cause. But insofar as this is true in Kant's logic, the notion of cause is the purely formal one of ground to consequent, where the latter is in some way explained or justified by the former. Yet Kant then goes on to interpret this relation in the way schematized for naturalistic explanation, as that of efficient cause to its effect. This would be appropriate if the relation of desert was meant to function in our scientific explanations of nature. However, this cannot be the case, because for Kant virtue makes sense only in a context of freedom, which cannot be ascribed to anything so long as it is considered a naturalistic phenomenon. If so, then the relation to ground-to-consequent must remain a practical and a moral one: that is, what it means for someone to deserve happiness because of her virtue is that there are moral grounds for a rational agent to try to bestow happiness in this way.

Of course, ordinary human beings may never be in a good moral or epistemic position to apportion happiness to virtue to this way. Reflection of our limitations may then indeed lead us to the idea of God, as the only agent who would be properly entitled to give us what we deserve. Even so, none of this would require any kind of belief that such a God exists. All that morality requires is that we recognize that a perfect world would be 4 a world in which God ensures that happiness corresponds to merit. Morally decent agents might then wish for such a world without willing that it come into being, insofar as they lack confidence in the existence of God or the immorality of the soul. The Highest Good would then not be a goal to be produced, but merely a standard for assessing the moral adequacy of the world.

Although Kant never explicitly repudiates the argument of the second Critique, there is no trace of it in his subsequent works, including Religion within the Limits of Reason Alone. Instead, in a long footnote to the introduction of the Religion, Kant makes it clear that the moral law does not need the Highest Good for either the authority or motivational power that Kant associates with that law. That the good may suffer and the wicked prosper is a possibility that we may just have to make our peace with:

> But that every human being ought to make the highest possible good in the world his own ultimate end…is a proposition that exceeds the concept of the duties in this world, and adds a consequence (an effect) of these duties that is not contained

> in the moral laws and cannot, therefore, be, be evolved out of them analytically. For these laws command absolutely, whatever their consequences; indeed, the even require that we abstract from such consequences entirely whenever a particular action is concerned, and thereby they make of duty an object of the highest respect, without proposing to us, or assigning, an end (and an ultimate end) such as would constitute some sort of inducement for it and an incentive to the fulfillment of our duty. All human beings could sufficiently partake of this incentive too if they just adhered (as they should) to the rule of pure reason in the law. What need have they to know of the outcome of their doing and nondoings in that the world's course will bring about: It suffices for them that they do their duty, even if everything were to end with life in this world, and in this life too happiness and desert perhaps never converge. (6:6n. My emphasis)

However, Kant does not then just dismiss the Highest Good as an instance where practical reason has reached beyond its own proper boundaries, even though doing so 5 would count as dialectical illusion of practical reason that would nicely complement the dialectic of practical reason. Instead Kant continues:

> Yet it is one of the inescapable limitations of human beings and of their practical faculty of reason…to be concerned in every action with its result, seeking something in it that might serve them as an end and even prove the purity of their intention—which result would indeed come last in practice (nexu effectivo) but first in representation and intention (nexu finali). Now, in this end human beings seek something that they can love, even though it is being proposed to them through reason alone….
>
> ….That is, the proposition Make the highest possible good in this world your own ultimate end,"…is introduced by the moral law itself, and yet through it practical reason reaches beyond the law. And this is possible because the moral law is taken with reference to the characteristic, natural to the human being, of having to consider in every action, beside the law, also an end (this characteristic of the human being makes him an object of experience). (6:6n, my emphasis)

Here Kant tell us that even though commitment to the Highest Good goes beyond the law's demands, such a commitment is nevertheless necessary for us because the human will, as an object of experience, requires some sort of end.

What's puzzling about this claim is morality seems to have already supplied the human will with all the ends it might need. If an end is just a goal or state of affairs to be realized, then we can rest assured that every morally commanded act also gives us a more specific end, such as repaying a debt, telling the truth, or easing someone's suffering. In addition to such particular ends, the moral law also gives us two very general ends that can never be discharged—the duties of beneficence and self-improvement, by which we are commanded to advance the permissible happiness of others and to perfect ourselves in both our moral and non-moral capacities. Admittedly, we could understand "end" in a still more abstract sense, as some material concern that is at stake in all our moral obligations, some common point by which they all deserve to count as moral concerns. 6 Yet even if we need such an end, it has already been supplied by Kant's conception of humanity. In the Groundwork Kant tells us that the morality requires a distinctive

"matter" or "end-in-itself", which he equates with humanity, the effective exercise of the powers of rational self-governance. This is the underlying point of moral action, to properly respect and value our common ability to lead lives of our own rational choosing, and so to be in some deep way responsible for ourselves. Our will's need for a moral end should be accommodated somewhere between the particular goals morality assigns us that this fundamental concern that animates and unifies them all.

However, in the footnote at 6:6 Kant does not merely say that moral willing requires some object, or even that it requires an object that can arouse our inclination. Instead, Kant says that the moral will requires as object of love. Now, by 'love' Kant might just mean any sort of desire, in which case it seems that we already have more than enough candidates for this sort of object. Elsewhere, Kant makes a distinction between "practical" and "pathological" love; practical love turns out to be morally motivatedbeneficence, while pathological love is merely a kind of strong fancy or fondness for something. If Kant is understanding love in either of these two senses, then there is no real problem here for the Highest Good to solve.

But here Kant may mean by 'love' is something more like what we would normally associate with the word, a deep ardor and devotion that can play a central organizing role among one's concerns. Such love would not merely be a kind of liking or fancy, but an instance of the special category of inclinations that Kant calls "passion" (Leidenschaft), which for Kant includes such "diabolical" vices as envy, vindictiveness, and malice. Kant defines inclination in general as "habitual sensuous desire"; the objects of 7 inclination are things that we take pleasure in, and the reason we care about them as we do is ultimately because we are pleased by their existence or displeased by their absence. What distinguishes the passions from other inclination is that the passions, although ultimately forms of self-love, nevertheless wear the guise of reason. Kant tells us tat the passions do not merely present some concerns to be weighed up against others; instead, they pretend to be authoritative interests that outrank those others.

For Kant, ordinary inclinations are only as strong as their affects; that is, the felt experience or pleasure or distress at the thought of their objects. The temptation posed by such inclination is weakened by reflection. In contrast, the passions, which pretend to provide rational standards, cannot only withstand but are strengthened by reflection, as when envy or resentment grow in power as we brood upon them. Kant likens the temptation that ordinary inclinations generate to the pressure that waters may put upon the damn that holds them back. Passion, in contrast, works like waters slowly eating away at and undermining the base of the dam.

Kant tells us that while ordinary inclination can be directed toward any sort of object, the primary object of every passion is some other agent. Passion is not just an impulse to do or consume something; rather, passion like reason claims presents its object not just as something pleasing, but as that to which I am in some way entitled. Kant tells us that every passion involves an element of "illusion" or self-deception whereby we convince ourselves that some subjective concern, grounded merely in the experience or anticipation of the agreeable, in really something we are owed by right. We normally desire admiration and esteem; we are guilty of the passion of envy when we think that others do us a wrong by looking better than us. We properly desire independence; we 8 betray the passion of ingratitude when we resent the aid that we come to need from

others. In these cases, some form of self-assertion cloaks itself in some rational concern. The passions for wealth, reputation, and power are all ways in which some form of selflove hijacks the norms of prudential reason (so that we go from thinking that money is always useful to the miser's view that money should always be acquired and never spent). Envy, ingratitude, and vindictiveness are all ways in which self-love manages to express itself in the language of justice.

Kant considers the passions to be much more dangerous than the ordinary inclinations. The ordinary inclinations do serve as temptations to wrongdoing, as so as occasions of "frailty" or weakness of will. Kant considers such failings to be merely episodic, however: we know we are acting weak-willedly when we do so, and such acts typically arose such self-corrective emotions as remorse and regret, that lead us to strengthen our resolve to do better in similar situations in the future. Kant tells us that despite these risks, the

> natural inclinations as good, i.e., not reprehensible, and to want to extirpate them would not only be futile but harmful and blameworthy as well; we must rather only curb them, so that they will not wear each other out but will instead be harmonized into a whole called happiness." (6:58).

The passions, in contrast, are far more insidious. Because the these artificial inclinations involve ways by which self-love passes itself off as reason, we may fail to realize that we are in their grip, and instead take pride in our supposed virtue. Since the passions involve a kind of self-deception about morality, they can quietly corrupt our entire practical outlook, contaminating the very capacities we need to recognize and combat them. Kant concludes that although it would be wrong to try to extirpate all the inclinations, we should strive to rid ourselves of the passions as much as we can. This 9 requires virtue not in the sense of self-control, but as "apathy", the ability to not listen to the blandishments of the passions. In this condemnation, Kant makes no exception for any kind of love. Insofar as love for a person, group, or tradition might present itself as a rival authority to morality, it too is something we must learn to be deaf to.

Yet despite their faults, the passions play an essential role in human moral development. Kant understands that development in terms of the unfolding of three basic "predispositions" or aptitudes, each which involves the way a certain kind of self-love in integrated with the authority of reason. The first predisposition is animality, the sort of natural teleology of our drives and affects insofar as they can be understood being directed toward our good as living organism. The second predisposition is humanity, where we become capable of articulating and pursuing a conception of happiness in response to background prudential and social norms. The highest disposition is that of personality, in which we become capable of taking a kind of moral satisfaction in ourselves in response to the moral law. What is important for Kant is that these dispositions do not emerge together. Rather, we realize animality first, and only from this does humanity and eventually personality emerge. Before a new form of practical reason can appear, the old form must first undergo a kind of distortion. It is in the transitions between predispositions that the passions can emerge, where the ways of thinking of one stage are disrupted but not yet fully reorganized by the norms of a higher stage.

If this is true, then the human being cannot avoid the rule of the passions. Instead, she must work through this rule, ultimately becoming liberated into true autonomy with the full emergence of the predisposition to personality. Unfortunately, Kant tells us that this process cannot be completed in any finite span of time. Although we are supposedly 10 morally obligated to become holy (or morally perfect), we can only hope to approach that status ever closer, as a kind of asymptote that we can never reach. Yet Kant does not conclude that the quest for moral perfection is futile. Instead, he claims that the entire, infinite approach to holiness can count as its attainment, when the series is taken as a totality in the mind of God.

True virtue, Kant holds, has the character of a timeless "revolution of heart," where we decisively subordination the maxim of self-love to the moral maxim in our way of thinking. This is what Kant calls the "virtus noumenon", which necessarily occurs outside of time (that is, it is not a datable event in our lives, but something more like the entirely of our lives taken as a whole). However, we are creatures who only know themselves as they are in time, in terms of a sequence of stages in which the later ones are determined by the earlier. From this point of view, the virtus noumenon expresses itself as the virtus phenomenon, the gradual growth in our capacities of moral self-possession. The virtus noumenon is an all-or-nothing thing; the virtus phenomenon is at best the steady but always incomplete detachment from the life of the passions.

Now here is the paradox. On the one hand, all the passions are bad, and true autonomy can only be achieved once we have liberated ourselves from their rule. On the other hand, the only way pure practical reason can appear to us in time is as some course of the passions. If so, then there must be some passions that, defective though they must seem, nevertheless count as the empirical manifestation of the noumenal power of reason. This is the sense, I propose, that morality in us needs an object of love. This is not a problem of practical reason as such, since it would not affect agents such a God or the angels, who do not need to come to reason through a process of self-development. 11 Instead, this problem only affects creatures like us who have to grow into their autonomy, and who can only know themselves as objects of experience in time. For such creatures, there must be a kind of passionate devotion that counts both as a kind of moral distortion, yet also as an incipient empirical manifestation of true reason.

How does commitment to the Highest Good play this role? All passions, as inclinations, require some kind of end that one can take pleasure in. The Highest Good provides such an end, where the associated pleasure is necessarily moral in character. The joy we take in the Highest Good derives in part from the moral ends of attaining both virtue and happiness, insofar as it is permissible. Yet the Highest Good adds a further source of pleasure that comes from the way these moral ends are combined. The attainment of virtue and happiness pleases the morally good person, but even more satisfying is the prospect of people getting what they deserve, simply because it is what they deserve. The Highest Good, by virtue of both its content and its form, should be supremely agreeable to any morally half-decent person.

As a passion, proper love would have to involve not just an object that we enjoy, but also some sort of claim with address to others. In the case of the morally deleterious passions, this claim is often a sense of entitlement, as a distorted version of the moral ideals of freedom and equal rights. In our devotion to the Highest Good, we encounter another way

in which as passion might make some kind of claim on others. Kant holds that we should conceive of the Highest Good as a collective task that falls upon humankind as a whole. If so, then the sort of expectations that follow are like those that come with any kind of joint activity, where each member can call upon her fellows to step up and do their share, just as she realizes that other can call on her to do her part. 12 The sense of entitlement associated with the Highest Good is thus the beginning of a real sense of being morally accountable to others.

Finally, the Highest Good has the special virtue of being a moral goal that we cannot discharge. The object of moral love needs to be such that it can sustain the right kind of passionate devotion to efforts to attain true autonomy, a task that Kant insists cannot be completed in any finite span of time. Other laudable goals, such as the production of an ideally just state, are such that they could, in principle, be realized at some point in time, so that they could no longer inform our moral efforts which must continue on without end. In contrast, we can never hope to fully bring about the Highest Good at any point in time; in part because we will never attain perfect virtue, but also because we could never gain complete mastery over the forces of nature so as to make the connection between virtue and happiness completely necessary. We will always experience our strivings to produce the Highest Good as incomplete, just like our efforts to attain true autonomy. Yet with God and immortality, we can at least retain the hope that we can always come closer and closer to it.

Hoping for Peace

Lee-Ann Chae

Abstract: How should the ideal of peace bear on practical reasoning in our non-ideal world? In the just war tradition, the ideal of peace does not feature very prominently, perhaps because it seems exceedingly improbable that we might ever achieve world peace. Just war theory aims at the more modest goal of creating a less violent world by exerting moral pressure on the practice of war.

I argue, however, that we should hope for peace. The account of hope I offer rejects traditional accounts that analyze hope into component parts of belief and desire. On the view I defend, hoping is an exercise of our agency that not only shapes the scope of our activities, but also gives our activities a meaning they would not otherwise have had. In acting on our hope, we reach out towards a possible future, and draw the value of that possible (peaceful) future into what we are doing now.

Introduction

How should the ideal of peace bear on practical reasoning in our non-ideal world? In the just war tradition, the ideal of peace doesn't feature very prominently. One reason for this neglect might be that given the world as it is, it seems exceedingly improbable that we might ever achieve world peace. And so, rather than aim at a utopian ideal, just war theory aims at a more modest goal – to create a less violent world by exerting moral and legal pressure on the practice of war. Hopes for peace are perhaps too akin to wishful thinking.

Against such a view, however, I argue that we should hope for peace. I argue that hope is a kind of orientation towards the future, towards others, and towards oneself, that makes it possible for us, as finite and so non-omniscient beings, to live together as we should – as moral agents, free from violence and coercion. The account of hope I offer rejects traditional accounts that analyze hope into component parts of belief and desire. On the view I defend, hoping is an exercise of our agency that not only shapes the scope of our activities, but also gives our activities a meaning they wouldn't otherwise have had. In acting on our hope, we reach out towards a possible future, and draw the value of that possible future into what we're doing now. And so when we act on our hope for a peaceful world, we understand our action as a (non-instrumental) part of humanity's coming to live in peace.

I. Traditional Accounts of Hope

Whether philosophers argue that hope is an emotion, a disposition, or a special kind of cognition, there seems to be a general consensus that hope can be reduced down to something that includes some kind of belief and some kind of desire. The belief at

issue involves the hopeful person's calculation of likelihood of attaining the hoped for thing. In order for the hopeful person to hope that *P,* she has to believe that the likelihood of attaining *P* falls somewhere between impossible and assured. If she believes that *P* is impossible, then her seeming hope for *P* is actually just wishful or magical thinking. One can't hope to be an elephant or to turn back time. If she believes that *P* is a future event that is certain, then she's not hoping for *P* as much as she is waiting, or planning, or looking forward to it. One can't hope that the sun will rise tomorrow or that Starbucks will have coffee. The issue of where, more precisely, within this spectrum my calculation has to fall is a matter of less consensus, with some philosophers taking a rather expansive view (that there is uncertainty as to whether the hoped for thing will happen)[45], some taking a slightly more restrictive view (that the hoped for thing be seen as possible)[46], and others taking a rather restrictive view (that the hoped for thing be seen as likely).[47]

Aquinas argues, "Hope is a movement of appetite aroused by the perception of what is agreeable, future, arduous, and possible of attainment."[48] If the hoped for thing were not arduous, we wouldn't need hope, since we could just work towards the end. If the hoped for thing were impossible, hope would be pointless. To hope well, in Aquinas' sense, we must be able to realistically assess our chances of attaining *P*.

Usually when we think about hopes, we think about the future, but the uncertainty involved doesn't have to be in the future. It could be in the past, and so settled. But from the hopeful person's subjective point of view, it is unknown or unknowable, and so uncertain. E.g., "I hope she got home safely last night" or "I hope he enjoyed his birthday." In order for us to hope for *P*, *P* doesn't actually have to be uncertain, just uncertain for us given the evidence available to us.

So how does this belief in probability feature in what I will call "traditional" understandings of hope? According to Luc Bovens, "Hoping *is* just having the proper belief and desire in conjunction with being engaged to some degree in mental imaging,"[49] where mental imaging consists in the "devotion of mental energy to what it would be like if some projected state of the world were to materialize."[50]

Bovens warns that our hopes should be clear-eyed in the sense that our beliefs about the probability of the hoped for thing attaining should be properly tied to the

---

[45] *See, e.g.,* David Hume, Book II, Part III, Section IX in *A Treatise of Human Nature*, eds. David Fate Norton, Mary J. Norton (Oxford: Oxford University Press, 2000); Luc Bovens, "The Value of Hope," *Philosophy and Phenomenological Research* 59, no. 3 (1999): 667-681.

[46] *See, e.g.,* R.S. Downie, "Hope," *Philosophy and Phenomenological Research* 24, no. 2 (1963): 248-251; John Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983), p. 32, explaining that A hopes that *P* when "(1) A does not believe that *P*; (2) A does not believe that not-*P*; (3) A believes that *P* is possible; (4) A desires that *P*."

[47] Hobbes and Day require that *P* is not only possible, but probable. For Hobbes, the end has to be seen as obtainable, or as he explains it, "Appetite with an expectation of success is called HOPE."[47] *Leviathan*, Book I, Chapter 6. For J.P. Day, "A hopes that P entails (1) "A wishes in some degree that P" and (2) "A thinks that P is in some degree probable" and "[t]hese two tests or conditions of the truth of "A hopes that P" are severally necessary and, it is submitted, jointly sufficient." Day, "Hope," *American Philosophical Quarterly* 6, no. 2 (1969): 98.

[48] Summa, Vol. 3, II(2), Question 17, First Article.

[49] Bovens, "The Value of Hope," p. 674.

[50] *Id.*

evidence, or else we risk slipping into wishful thinking. When we wish, we raise the subjective probability of the wished for thing beyond what is warranted by the evidence. The line between hoping and wishing is difficult to guard because wishful thinking is so seductive, and this is what makes hoping so dangerous.

Philip Pettit takes a different tact, and puts our beliefs about the probability of the hoped for thing attaining on the outside of hope – we still make these probabilistic calculations, and while our hope is responsive to these calculations, the calculations are not strictly speaking a component of our hope. When we hope, we put our actual belief about probability "offline;" we are moved to act as if the hoped for end were going to attain (or at least as if there were a good chance).

Even though we act as if things were otherwise than we believe, one reason why hope is pragmatically rational, according to Pettit, is that it lifts us out of panic or depression and gives us control and direction.[51] Hoping protects us against emotional collapse and a loss of self-efficacy when the chances are especially low.[52] In the face of such trying odds, hope gives us a way to hold ourselves up and to keep going on.

It seems true enough that in some instances, we might use hope as a kind of shield against low odds, because otherwise there would be only despair.[53] But I don't think this is the best or only way to understand hope. When hope's rationality depends so heavily on its instrumental value in helping us to attain our hoped for end, we might end up with the following result, which, if you're like me, will make you uneasy. Consider two young students, one who goes to a terrible school, and one who goes to a terrific school, and what their hopes for a bright future look like. If hope is a shield, we might be led to say that the student who goes to the terrible school should hope more than the student who goes to the terrific school for a bright future, because of the longer odds. (Or, even worse, on a Bovens-like analysis, we might be led to say that the student who goes to the terrible school should hope less than the student who goes to the terrific school, again on account of the long odds.)

Analyses of hope that focus so narrowly on beliefs about the probability of attaining the hoped for end suffer from two main difficulties. First, they do not adequately distinguish between hoping and wishing, or hoping and trying. If, on the one hand, hoping is a kind of irrationality, in that we set aside what we know to be true, it becomes harder to distinguish it from mere wishing. And to say that the difference between hoping and wishing comes down to a miscalculation of the odds of attaining the desired end makes hoping too much of a kind with wishing. And if, on the other hand, hoping is a kind of prediction of success, it becomes harder to distinguish it from trying. And second, these belief-based approaches, in focusing so exclusively on how the world determines or shapes our hopes, loses sight of the sense in which hope is something we bring *into* the world. A compelling account of hope should be able to explain how our hopes motivate us and give meaning to our lives. I don't think the reason why we have

---

[51] Philip Pettit, "Hope and Its Place in Mind," *The Annals of the American Academy of Political and Social Science* 592, no. 1 (2004): 152-165, p. 161.

[52] *Id.,* p. 157.

[53] What is the opposite of hope? According to Hume and Bar-Tal: fear. Day: fear, resignation, despair, and desperation. Ratcliffe: depression, loss of aspiring hope, demoralization, loss of trust. Govier: despair, cynicism, fear, pessimism.

certain hopes and not others could boil down to (something that includes) our beliefs about the probability of attaining the hoped for end. And so instead of theorizing what kind of calculative belief is involved in hope, I'll follow Margaret Urban Walker and Victoria McGeer in treating hope as a primitive.

According to Walker, hope is "a recognizable syndrome" that cannot be identified with "a single 'recipe' of specific ingredients in precise proportions." Rather, we should recognize that "there are patterns of ingredient perceptions, expressions, feelings, and dispositions to think, feel, and act that are part of the repertory of hopefulness."[54] McGeer picks up on Pettit's connection between hoping and agency, but for her, the connection is much tighter. Rather than seeing hope as something that protects our agency, McGeer sees hoping as a way of exercising our agency. So when we hope in the face of long odds, "our persisting capacity to hope signifies that we are still taking an agential interest in the world, and in the opportunities it may afford, come what may."[55] To hope involves recognizing, but not feeling constrained by, our limitations as finite human beings. To hope is to learn, to be creative, and to be energized in the face of those limitations and sometimes to push beyond them.

I'll build on Walker's and McGeer's accounts because I think there's more that can be said about the structure of hope. I'll offer a preliminary account of hope that can explain the role it plays in motivating our actions, and in giving meaning to our activities and experiences.

II. Meaningful Hope

In trying to understand the value of hope, I'd like to begin by considering what it's like to live without hope and why such hopelessness is bad. Descriptions of hopelessness often share two elements. First, one who is hopeless cannot see a future for herself; she cannot imagine a future and a see a place for herself in it. When there is no future horizon that calls, what's missing is not only the lack of direction, but also the feeling that something different is possible. And second, one who is hopeless forgets that things were not always so; she forgets how things used to be. She cannot remember that things used to be different than they are now. With no light ahead, and no memories behind, the person living in hopelessness is entombed in the present.

Then what makes hopelessness so bad isn't just that hopelessness makes it less likely that you'll attain some particular end, or that you'll become efficaciousy inert. What makes hopelessness so bad is that it confines you to the present bad moment, to a moment that has no meaning that relates you to a different and brighter future. If this is what makes hopelessness so bad, it gives us a clue as to hope's value.

We can contrast a life lived without hope to a life lived with hope. When we hope, the time horizon expands out from the now and we see different possible futures.

---

[54] Margaret Urban Walker, *Moral Repair: Reconstructing Moral Relations After Wrongdoing* (Cambridge: Cambridge University Press, 2006), p. 48.

[55] Victoria McGeer, "Trust, Hope and Empowerment," *Australasian Journal of Philosophy* 86, no. 2 (2008): 237-254, p. 246.

In acting on our hope, we not only reach out towards a possible future, but we also draw the value of that possible good future into what we're doing now. When we act on hope, we see our hopeful action as a moment that could be a part of the hoped for end, and so it has a different meaning for us. We see our hopeful actions as meaningful because they are an early part of realizing the future good.

Hope can serve as a rational ground for action that doesn't just reduce to an instrumental trying. In thinking about whether to try to accomplish some end, the rationality of trying can depend on the belief that the particular trying action has a good chance of contributing to bringing about the end. If a friend were considering trying to undertake some activity where the chances of success were very low, we might advise her not to even bother trying. If the chances are very low, it might be irrational to try. So, for example, it would be irrational to try to win the lottery by buying extra tickets, or to try to build a house with no knowledge of carpentry, or to try to learn a foreign language in a week. It would be hard to find any sense in those activities as a trying to bring about some end.

Compare the lottery ticket buyer, or the would-be house builder, or foreign language learner, on the one hand, to a protester at a peace march who is opposing her government's war posture or imminent prosecution of a defensive war, on the other. What is she doing there? Seen as a trying, we can understand why a reporter might ask a peace protester why she bothers to march. It's hard to see how waving a banner could prevent a bomb from being dropped, or the chanting of a slogan bring about a cease-fire. Protesting is not a sensible way to try and end war. And when pressed by the reporter what she thinks the chances are that her participation in this march increases the chances of ending the war, she might give an answer like "almost none" or "very low," increasing the reporter's befuddlement.

Traditional accounts of hope are not sure what to make of hopes for world peace, either. I suspect that traditional hope theorists do not find hopes for peace sensible because traditional hope understands hopeful actions as a trying. Bovens briefly discusses hopes for peace in a footnote. As he explains it, when I hope for world peace, either (a) "the projected state in utopian hopes functions as a guiding ideal," in which case "what I am hoping for strictly speaking is that the world will move closer toward peace in my life time and it is not true that I am confident that *that* will not come about" (*i.e.*, I'm not confident that it *won't* happen)[56] or (b) I have a divided mind – I admit that according to the evidence, I should be confident that world peace *won't* come about in my lifetime, but part of me resists this confidence, which enables me to continue to hope.[57] And Pettit uses the prevention of war as an example of something that a potentially hopeful agent cannot influence. If we are to believe that the prospect of a war's not taking place is beyond our influence, and makes trying insensible, what are we to make of the following case?

---

[56] I find this option uncompelling because it cannot explain why uncertainty in this case would lead the protester to march. Afterall, we're uncertain about many things that don't lead us to action. I would be uncertain crossing the street without looking both ways that a car wouldn't hit me, but I'm not going to cross the street without looking on that basis.

[57] Bovens, n. 4. On Bovens' own account, such a "hope" would actually constitute a kind of wishful thinking.

During the Bosnian War, on May 29, 1992, at 4 p.m., Vedran Smailovic witnessed the obliteration of 22 people who had been queuing at a bakery in Sarajevo. We would've been able to understand if Smailovic had been driven to hopelessness in the face of such inhumanity. But he wasn't. "I am nothing special, I am a musician, I am part of the town. Like everyone else, I do what I can."[58] Here's what Smailovic, a concert cellist, decided he could do. Every day at 4 p.m., he put on his full concert dress, took his cello to the site of the bread massacre, and played Albinoni's Adagio in g minor. As civilians dodged sniper fire and took cover from Serbian bombs, Smailovic played out in the open for 22 days.[59] He also played in cemeteries, flooded with makeshift graves, which was especially dangerous because snipers would pick off civilians who came to mourn or bury their dead. As a self-avowed pacifist, Smailovic became a symbol of civil resistance during the war by playing his cello to "daily offer a musical prayer for peace."[60]

When a reporter asked him if he wasn't crazy for playing his cello while Sarajevo was being shelled, Smailovic replied, "You ask me am I crazy for playing the cello, why do you not ask if they are not crazy for shelling Sarajevo?"

So who was right, the reporter or Smailovic? If we don't think he was crazy, there must be something more to protesting than merely trying to cause an end to the war (for surely it seems insensible that one would try and cause the end of war by playing the cello). And that is, trying is not the only practical stance towards a possible future that helps to rationalize or motivate present action.

Take the familiar example of spending time with someone in the hopes of getting to know her better. It would be a mistake to think of my activities with her as having merely instrumental value, in that they increase the probability of attaining my hoped for end of friendship. It might be true that my activities do in fact have the effect of increasing the probability, but that cannot be my reason for doing them.

Rather than thinking of my actions as instrumental tryings aimed at the attainment of my hoped for end, it's better to see how it is that my actions are informed by my hope. Because of my hoped for end, I undertake certain activities with my potential friend that I wouldn't otherwise have done – we listen to music together, go for hikes, watch each other's dogs. But not only does my hoped for end guide the scope of my activities, it also gives my activities a meaning they wouldn't otherwise have had. Because I see myself as in the process of constructing a friendship, my interactions are characterized by an attitude of openness and curiosity, and I am oriented to my potential friend as a whole person.

But just as I shouldn't see my activities with my potential friend as mere means to some end, or as merely increasing the probability of my hoped for end, I shouldn't see

---

[58] http://www.nytimes.com/1992/06/08/world/death-city-elegy-for-sarajevo-special-report-people-under-artillery-fire-manage.html?pagewanted=all

[59] Here's another place where a Pettit-style analysis does not seem to go far enough. To say that Smailovic was acting as if things were otherwise than they were is to lose sight of the courage it took for him to play. His act was courageous because it was dangerous, and he knew it was dangerous – he was not just making-believe that everything would be alright if he could play his cello.

[60] I am not taking this quote literally. But it would be interesting to consider the question of whether petitionary prayer counts as hope on my account.

them as merely isolated incidents, either. I am not just in the moment of each activity, so to speak, and I do not greet her day after day with surprise: *Oh, there you are again!* Rather, the various activities we undertake together are held together by the value of the hoped for end. And that is because our actions pull the value of the hoped for end into what we're doing now. If I were not acting on the hope of getting to know you better, this hike we're taking together now would have a different character and a different meaning than it does in fact for me. And so if and when we do become friends, I won't be able to point to a specific moment we became friends, but I will be able to point to a history together which will have the character of a friendship in blossom.

Compare the hopefulness of getting to know someone to the hopelessness of getting to know someone. Let's say I'm meeting a famous poet at a reading. In instances like these, our social roles, which are supposed to help us navigate the world, end up limiting us in ways that can be frustrating. I might feel constrained in reaching out to the poet as a person who appreciates her work, and might feel I can only greet her as a fan. And if I do, our meeting will have a different character than if I have hope of getting to know her. It's true, of course, that perhaps the outcome will be the same whether I greet the poet with hope or with hopelessness, in that she and I don't become friends, but hoping is not outcome oriented in the same way that trying is.

There's a partial analogy here between meeting someone in the hopefulness of a friendship, and hope for peace. When faced with a violent aggressor, my nonviolence doesn't have to be an instrumental trying to bring about peace. When I can bring myself to hope for peace, I meet violence with nonviolence because it's possible that the members of the human community will live together peacefully, and I see my action as an early part of that possible peaceful world. I see this moment of nonviolence as a moment that could be a part of a peaceful world, and so it has a different meaning for me.

So there's another way to understand what the protester is doing, such that it would be difficult for the protester to make sense of the reporter's question: why bother protesting when there's no chance it will stop the war? Her trying to stop the war (if she's trying to do that at all) doesn't exhaust the value of what she's doing there because her actions are also marked by the value of the hoped for thing, peace. What the protester is doing is acting *on* her hope. This doesn't mean that she's there to buffer herself against cold, hard probability, and it doesn't mean she's there to stave off emotional and agential collapse. Nor does it mean that she's acting as if things were otherwise than the evidence suggests. Hoping is not a kind of irrationality.

Rather, the protester is looking out at the world through her hope. As the person who trusts sees the world through her trust, the person who hopes is guided by that hope in picking out what factors count as salient, in interpreting how they are salient, and in deciding how to act based on that interpretation. As someone who hopes for peace, she has become good at interpreting the world in ways that sustains her hope and orients her towards fulfilling it. This is why although we would advise our friend, in the face of low odds, not to bother trying to win the lottery, we cannot advise our friend, again in the face of low odds, not to bother hoping for peace.

So the connection between hoping and agency is stronger than: if I don't hope, I might lose agency (either in this endeavour, or some other). Hoping for peace *is* a way of

exercising our agency. What the protester is doing is living in the possibility of peace. She is in the process of constructing a reality she believes is actually possible.

The protester can act now, taking as her reason for action the possibility that her action forms a part of the eventual end that she seeks. To act on hope for peace is to be part of the movement that might end in a peaceful society. From the vantage point of the peaceful society, we will be able to look back on Smailovic's playing, and recognize it as an early part of the effort for peace.

Conclusion

What is it like to live without hope? Simone Weil explains the hopelessness of the soldier engaged in the Trojan War by explaining that death is the future his profession has assigned him.[61] For the soldier, every moment is essentially tied up with the possibility of death. And so "every morning, the soul castrates itself of aspiration, for thought cannot journey through time without meeting death on the way."[62] Permeated with death, the soldier is confined to live moment to moment. And who, in a moment where she finds herself confronted by an armed enemy, can give up the sword?[63] So the killing goes on, because without hope, there is no way out.

To think about a peaceful future together is not wishful thinking, as some might warn. Wishful thinking is a kind of escapism, indulging in the pleasure of wondering: *what would it be like if…?* But to hope for a peaceful world isn't just to indulge. When we hope for peace, we understand our current actions as meaningful contributions to peace, and we prepare ourselves – morally and materially – for the hoped-for eventuality, so that instead of just being people who say we're for peace, we can become the kinds of people who are capable of it.

[61] Simone Weil, "The Iliad, or the Poem of Force," Chicago Review 18, no. 2 (1965): 5-30, p. 19.

[62] *Id.*

[63] *Id.*

Biographical Identity and Retrospective Attitudes

Camil Golub (camil.golub@nyu.edu)

[Draft for NUSTEP 2017—please do not cite without permission]

Abstract: We all could have had better lives, yet often do not wish that our lives had gone differently, especially when we contemplate alternatives that vastly diverge from our actual life course. In this paper I ask what, if anything, accounts for such conservative attitudes. First I examine some possible answers: (i) the lack of direct psychological connections with our merely possible selves; (ii) a general conservatism about value; (iii) the importance of our actual relationships and long-term projects. I argue that these answers are all incorrect or incomplete. Then I offer my own proposal, inspired by R.M. Adams' (1979) answer to the problem of evil: it is reasonable not to regret many things in our past because they contributed to who we are. Our biographical identities constrain the live options for our retrospective attitudes. I end by connecting this proposal with recent narrative accounts of personal identity.

1 The puzzle

We routinely make judgments about how good our lives are, or could have been. And when we judge that a certain life would have been better or worse for us, this usually supports retrospective attitudes like regret and affirmation.[64]

Sometimes, however, we judge that certain lives would have been better for us, all things considered, and yet do not regret having missed out on them. Indeed, we affirm our actual lives when comparing them to those better alternatives. Here is an example:

FRANCE Suppose that I justifiably believe that, if my parents had emigrated to France when I was a child, my life would have been better, according to my actual standards for a good life. Nevertheless, I do not regret having missed out on this better life.

Note that FRANCE cannot be diagnosed as an instance of the familiar conflict between moral concerns and self-interest or personal value: the setup is not that, had my parents emigrated to France years ago, the world would have been *morally* better, or better from an impartial point of view.[65] Rather, the tension arises between self-regarding

64 I understand regret and affirmation as retrospective preferences: for instance, to regret that one did not go on vacation last summer is to wish that one had gone on vacation. I will have little to say about the emotions, e.g. bitterness or nostalgia, that often accompany such retrospective preferences.

65 Compare with the question discussed by Wallace (2013): how can we affirm the value of many things in our lives, and indeed our lives as such, given that this seems to entail affirming events that made the world objectively worse? If, say, the Rwandan genocide had not taken place, many events that have shaped my life wouldn't have happened either: for example, I probably would not have met my spouse. And if the Holocaust had never happened, then I would not have existed. If I believe that it is overall better that I did meet my spouse or that I exist, then it looks like I am committed to affirming features of the world that are objectively lamentable. When addressing this issue, the following diagnosis seems plausible: from a moral standpoint, I should recognize that it would have been better for the Rwandan genocide and thus for

retrospective attitudes: I judge that a certain life would have been better *for me*, and yet I affirm my actual life course. How can we make sense of this?

This conservative bias is pervasive in our retrospective outlook. There are *many* better lives we could have had—e.g., lives in which we would have grown up in better neighborhoods, gone to better schools, or made wiser decisions in our youth—and which we nevertheless do not regret, even upon careful reflection.

Two features of these conservative retrospective attitudes stand out. First, it seems that we can affirm things of *disvalue* in our past, e.g. experiences of adversity and hardship, and not just things that are *less valuable* than what we could have had. In some cases, we might affirm intrinsically disvaluable things because we find some instrumental value in them: for instance, painful experiences can teach us something important, facilitate valuable relationships, or build our resilience and integrity in the face of adversity. But this is not always the case. When I think of disvaluable experiences in my past, and compare them to better lives from which they would have been absent, my attitude of affirming those experiences does not seem to be primarily grounded in their instrumental value. A different kind of attachment is in play.

Secondly, the conservative bias in our retrospective attitudes gets stronger the more distant we are in time from the events that could have brought us a better life. If I judge that, had something happened *yesterday*, I would have had a better life, this gives me much stronger reasons for regret than the better life I could have had in FRANCE. Any good theory of these matters should explain why regret recedes over time in this way.

Before examining in detail some potential justifications for conservative retrospective attitudes, let me briefly put aside other possible responses to cases like FRANCE.

Someone might suggest that my lack of regret for the better life I could have had is desirable, because regret is generally harmful to oneself. Or that it is pointless to regret what might have been, because we cannot do anything about the past. This may all be true. But the question I am interested in is whether regret is nevertheless a well-grounded or fitting response in cases like FRANCE, putting aside any prudential considerations that might count against it.[66]

A different response would go as follows: we often do not regret what might have been because we cannot really *know* that we would have been better off had things gone differently. Perhaps my lack of regret for the life I would have had in FRANCE similarly comes from my deep uncertainty about what that life would have looked like. But let us assume away such epistemic obstacles. Even if the value of my possible life in FRANCE is questionable, there are again many other possible lives that *would* have been better for

---

me to not to have met my spouse, or for the Holocaust not to happen and for me not to exist, but from a self-regarding point of view, it is rational not to regret my own existence or the things I value in my life. This diagnosis, however, is not available for cases like FRANCE. For more on the potential mismatch between retrospective moral judgments and regret, particularly in the context of the non-identity problem for reproductive choices, see Parfit (1984), Ch. 16.

66 Note, moreover, that it is typically the emotional states associated with regret, e.g. bitterness or anger, that have corrosive psychological effects. Mere retrospective preferences are arguably much less harmful.

me, and which I do not regret. Moreover, my lack of regret would remain reasonable, I believe, even if I pictured the good-making features of those lives in vivid detail. I believe the puzzle I am discussing has little to do with our non-ideal epistemic circumstances.[67]

Another option would be to say that it is rational not to regret a better life if one's actual life is *good enough*. We could call this a "sufficientarian" account. I am putting aside this response for two reasons. First, it does not account for any positive reasons we have for *affirming* our actual lives when comparing them to better possible ones, and I believe that we have such reasons. Secondly, in some cases regret does seem to be a fitting attitude, even if our actual lives are good enough by any reasonable standard. (Think of the moments after botching a job interview.) I am looking for an account that explains the difference between such cases and scenarios like FRANCE.

Someone might also redescribe FRANCE as a case where one's personal identity is at issue. For example, Leibniz thought that each of us only exists in one possible world. If this were true, then we could never regret the better lives that *we* could have had, because we only exist in the actual world.[68] And other, less radical theories of personal identity might deliver similar verdicts.

This response touches on something important: our self-conception influences indeed our conservative retrospective attitudes, and the proposal I will defend is an attempt to articulate this thought. However, I will take it for granted that our personal identity, understood in an austere metaphysical sense, is not at stake in cases like FRANCE. I find it highly natural to say that the Camil Golub whose parents emigrated to France is numerically identical to me.[69] When I find myself not regretting that I did not have his life, this is not grounded in any insight into our metaphysical separateness as persons, but in something else.

Finally, my conservative attitudes might simply be dismissed as instances of *status quo* bias. If we assume away all the pragmatic, epistemic and metaphysical issues mentioned above, someone might argue, it is irrational to affirm our actual lives when comparing them to lives that we think would have been better for us.

Now, I will not attempt to convince anyone of the rationality of the conservative attitudes I am describing. Rather, I want to invite those who share my judgments about cases like FRANCE to explore what would be the right account of such attitudes. But even

---

67 Setiya (2016) suggests a different epistemic diagnosis, centered on the idea of *specificity*: affirming lives we believe to be inferior is rational, not because we are uncertain about the value of the alternatives, but because of the *richly textured knowledge* we have of the valuable things that compose those inferior lives, compared to the abstract knowledge that some possible lives would have been better for us. I do not have the space to properly discuss Setiya's proposal here, but I'd be happy to talk about it in the Q&A.

68 "You will insist that you can complain, why didn't God give you more strength. I answer: if He had done that, you would not be you, for He would not have produced you but another creature." (Leibniz, *Textes inédits*, *apud* Adams 1979, p. 53)

69 Of course, others might not find this as natural as I do. I do not mean to beg any question here against metaphysical accounts of personal identity on which a numerically different person would have taken my place in FRANCE. Rather, my point is that, for those of us who think that we *are* numerically identical to some possible persons with vastly different life paths, lack of regret in cases like FRANCE cannot be explained by beliefs about metaphysical identity.

those who see these attitudes as irrational might find interest in this project: it could offer them an error *theory*—an explanation of why people like me have misguided attitudes about the past.

In what follows, I will look at three possible accounts of our conservative retrospective attitudes, and explain why I find them wanting. Then I will articulate my own proposal, centered on the notion of biographical identity.

2 No selfhood relations?

Velleman (2015) has a radical take on regret for what might have been. On his view, regret is not just inappropriate, but metaphysically confused in scenarios like FRANCE, because such cases do not allow for genuine self-concern. This is not because I am numerically different from the *person* Camil Golub whose parents emigrated to France. I *am* metaphysically identical to that person, Velleman would say, but he is not a *self of mine* in the sense that makes intelligible attitudes like regret and affirmation.

This argument relies on a conception of selfhood developed in Velleman (1996). On that view, self-to-self connections obtain just in case one can reflexively pick out in memory or anticipation a past or future self: for instance, I am on "first-personal" terms with my seven-year old self because I can refer to him and his experiences in an unmediated way simply by using the pronoun *I*.

Such first-personal connections are not possible, Velleman would argue, in cases like FRANCE. In going back to the common starting point of the two possible life paths and then up on the merely possible one, I lose the right kind of internal communication between selves. This rules out genuine first-person reference, and thus the intelligibility of first-personal attitudes like regret. Therefore, I should never regret not having what I could have had, because no self of mine could have had it.[70]

Now, let us put aside any worries we might have about Velleman's conception of selfhood, for instance whether it can circumscribe the right kind of causal and informational connection between selves in a substantive, non-circular way. For our current purposes, the main problem with this view is that it makes far too many cases of regret irrational or confused.

Take the following example:

ADMISSION Sonya applied for the PhD program in economics at Princeton, and has just learned that she was not accepted. She bitterly regrets that she did not get in.

Velleman would issue the same verdict here as in FRANCE, and for the same reason. Sonya's regret is metaphysically confused, because she cannot think of the possible Sonya that did get accepted to Princeton as a self of hers. She can perhaps *envy* this

70 Velleman (2015): "The person who might have been better off today if I had done differently in the past (...) is inaccessible to my self-concern. Of course, he is who I might have been—that is, who could have been a future self of my past self (...) But (...) selfhood is not transitive: another future self of my past self is not a self of mine. The fate of a merely possible self of mine is no more pertinent to me than anyone else's, since I can only imagine undergoing that fate." (p. 96)

merely possible Sonya, but only in the way that she can have such attitudes towards other people.

This is clearly wrong. In cases like ADMISSION, it is perfectly reasonable to regret the better lives that *we* could have had.

Velleman does have something more to say about such cases in which regret for what might have been seems rational. His proposal is that we can feel *vicarious* regret on behalf of our *past* selves, who were deprived of a better future.[71] According to Velleman, this diagnosis also explains the time-sensitivity of regret: as the distance in time between us and our past selves grows, we become more detached from their interests, and thus our reasons to feel vicarious regret on their behalf get weaker.

But this explanation does not do justice to the phenomenology of regret for what might have been or to the natural ways in which we articulate such regret: we identify with our merely possible selves, and wish that we were living their lives instead of our actual ones. The right account of these matters should allow that it is coherent and reasonable for us to experience regret in this way, without any circuitous reinterpretation of our attitudes. Moreover, it should allow that, even in a case like FRANCE, regret is intelligible and rationally permissible, although not a live option for many of us.

3 Particular value

The next option I will consider is based on G.A. Cohen's (2012) conservatism about value. Cohen's discussion focuses on prospective attitudes: his goal is to articulate the reasons we have for preserving actual valuable things at the expense of new and better ones. But his view easily extends to retrospective attitudes as well.

According to Cohen, one major source of support for conservative attitudes is our attachment to *particular value*—valuing something "as the particular valuable thing that it is, and not merely for the value that resides in it" (p. 148). If an intrinsically valuable thing actually exists, he says, this gives everyone reason to wish to see it preserved, at the expense of new and better things.[72]

This idea can also be applied to our retrospective attitudes. If Cohen is right, actual valuable things in our past give us special reasons to affirm their value, when compared with better but merely possible things. This might explain cases like FRANCE.

Moreover, Cohen's view allows that the normative force of particular value may be overridden if the difference in value between actual things and their alternatives is large enough.[73] This could explain the rational permissibility of regret for what might have been.

---

71 "When I complain, 'I could have been better-off,' I don't mean, 'I have a better-off possible self'; I mean, 'I (in the past) had the chance of being better off in the future'." (Velleman 2015, p. 96)

72 Another source of justification for conservative attitudes that Cohen discusses is *personal value*, which arises from personal attachments to valuable things. I discuss this option in the next section.

73 Cohen (2012): "Conservative conviction (...) exhibits a bias in favor of retaining what is of value, even in the face of replacing it with something of greater value (though not, therefore, in the face of replacing it with something of greater value no matter how much greater its value would be)." (p. 149)

Finally, Cohen's notion of particular value also captures the sense that, in cases like FRANCE, the explanation for our attitude of affirmation is not that we assign *more value* to our actual lives. Rather, we value our lives in a special way, when comparing them to lives that we acknowledge would have been *better*.

Despite these virtues, however, Cohen's view faces important problems in accounting for the conservative bias in our retrospective attitudes.

A first issue is that it seems reasonable to have robust conservative attitudes about the past—to affirm our actual lives when comparing them to many better lives we could have had—and yet be less conservative when it comes to preserving actual valuable things or replacing them with new and better things in the future. For instance, someone might be retrospectively attached to the city in which she has lived for a long time, and not wish that she had moved elsewhere years ago, and yet feel ready to move to a different stage in her life, including to a new city that better meets her needs and aspirations. Cohen's view seems unable to make sense of such a temporal asymmetry in our attitudes.

Secondly, particular value, as Cohen defines it, gives everyone equal warrant for conservative attitudes. But this does not seem right for the self-regarding attitudes we are interested in. It is implausible, for example, that everyone has equal reason to affirm *my* actual life when comparing it to the better life I could have had in FRANCE.

Finally, not all particular valuable things warrant a conservative bias. Cohen himself acknowledges this when discussing a counterexample to his view, proposed by David Wiggins. Think of an actual rosebush that has intrinsic aesthetic value, says Wiggins: there seems to be nothing wrong with replacing that rosebush with another rosebush that is qualitatively the same, putting aside any personal attachments we might have to it. In response to this challenge, Cohen concedes that perhaps only some things warrant a conservative bias, and notes that this concession invites "an interesting research program, into what forms of value demand preservation and what forms do not" (p. 165).

I agree that this research program is needed. And I believe it should cover our conservative retrospective attitudes as well: we need a deeper account of why such attitudes are warranted in certain cases, and with respect to certain things in our past. Merely appealing to the actuality of our life course is not enough.

4 Personal value

The third option I will examine is an account in terms of *personal value*. Some moral philosophers hold that we can reasonably value certain things in a privileged way because of the relations we bear to them: for instance, that it is permissible, and perhaps even obligatory, to care more about our own children than about other people's children.[74]

It is tempting to think that our conservative retrospective attitudes are explained by such personal relationships and attachments. For example, if my parents had moved to France when I was a child, I would have never met my spouse and wouldn't have made

74 See Scheffler (1997) and Kolodny (2010) for canonical treatments of the special reasons for desire and action that are provided by our relationships, long-term projects and other personal attachments.

the good friends that I have; I probably also would have ended up doing something other than philosophy for a living. Perhaps this is why I have special reason to affirm my actual life course, when comparing it to a life from which my important relationships and long-term projects would have been absent.

Such a view would also easily account for the difference between cases like FRANCE and ADMISSION. The relationships and projects I have developed since childhood warrant my lack of regret in FRANCE, it might be thought, while in ADMISSION Sonya presumably hasn't had time to develop any meaningful attachments since learning that she was not accepted to Princeton—attachments that would have been threatened by her being accepted. This might be why she has no reason to affirm her actual life course to the expense of the better alternative.

Unlike a particular value account, a personal value theory would not entail that everyone has equal reason to affirm an individual person's life course. For example, on this view, the fact that I would not have met my spouse had my parents emigrated to France years ago might give the reader *some* reason to retrospectively prefer that things went as they did, but not the same reasons that I have for this preference.

A personal value approach to our conservative retrospective attitudes, then, has considerable explanatory power. However, it too faces important challenges.

First, this account does not seem to leave room for a general asymmetry between our retrospective and prospective attitudes—a problem it shares with Cohen's conservatism about value. Someone could reasonably have strong conservative attitudes about her past, but a less conservative outlook on whether to preserve or privilege her current relationships and projects into the future. Think of someone who decides to end a long-term romantic relationship. Whatever her reasons might be for moving on, this person need not think that those reasons warrant *regret* about the years she has devoted to that relationship: she may wholeheartedly affirm her actual past when comparing it to a possible life in which she would have been romantically attached to a person she judges as a better fit for her. Making sense of such a psychological profile might not be an insurmountable challenge for personal value theories, but it does look like a difficult task.

Another problem is that it can be reasonable to affirm one's actual life course even when comparing it to better lives in which one's personal attachments would have been the same: think of a life in which you would have been involved in the same relationships and long-term projects, but some memorable moments in your past would have been replaced by *better* experiences. In this case too lack of regret and affirmation seem rationally permissible.

Perhaps the biggest problem for both personal value and particular value theories is that they cannot account for cases where we affirm things of disvalue in our past, and not merely things that are *less valuable* than what we could have had.[75] Again, we often find ourselves affirming retrospectively such intrinsically disvaluable things—for

75 On Kolodny's view, the agent-relative reasons provided by a relationship must *resonate* with the agent-neutral value of the discrete encounters composing that relationship. If the relevant encounters are objectively disvaluable, the relationship does not give rise to any positive reasons for partiality.

instance, experiences of adversity and hardship—and this need not always be due to their instrumental value.[76]

The proposal I will defend makes sense of these features of our attitudes about the past.

5 Biographical identity

A commitment to our biographical identity accounts, I believe, for the conservative bias in our self-regarding retrospective attitudes. Certain experiences, relationships, and projects in our past have shaped who we are, in a sense that is looser and thicker than bare metaphysical identity, and judgments about our identity thus understood can interact with, and be weighed against, judgments of value in guiding our retrospective attitudes.

This proposal is inspired by R.M. Adams' (1979) response to the problem of evil. After arguing that we should not be angry at God for the evils that preceded our existence, because we would not have existed in their absence, Adams also argues that we should not regret many evils that happened *after* our birth or even evils that are part of our own lives, for reasons that also concern our identity, but not in a metaphysical sense. Strictly speaking, we would still have existed in the absence of such evils, he says, but our lives are shaped by those evils so profoundly that wishing that they had not occurred would be close to wishing that someone else had existed instead of us.[77]

Now, I do not want to endorse the idea that it is, all things considered, rational to affirm our own actual lives when comparing them to possible worlds from which great evils would have been absent. Again, my topic is not the tension between personal value and moral concerns in our retrospective attitudes. Putting this issue aside, however, Adams' view suggests the right account of the conservative bias in our self-regarding retrospective attitudes: we can rationally affirm our actual lives when comparing them to lives that would have been *better for us*, due to a commitment to who we are, in a non-metaphysical sense. When we contemplate possible lives that would have been better for us but significantly different, and find ourselves affirming our actual life course, this is explained by an attitude of detachment from our merely possible selves: although numerically identical to us, they wouldn't have been *us*, in a different, ethically loaded sense.[78]

---

76 Kolodny (2010, p. 181) tries to accommodate such cases by appealing to the alleged agent-neutral importance of adversity. However, I find it implausible to assign any intrinsic agent-neutral value (or indeed, agent-relative value) to adversity or hardship as such. I suspect any lingering intuition to the contrary comes from the instrumental or enabling value of such experiences, e.g. their role in building our character, or in facilitating valuable relationships and projects.

77 Adams (1979): "Even if I could, metaphysically or logically, have existed without most past evils and their consequences in my experience, I doubt that that existence could have been *mine* in such a way as to matter much from the point of view of my self-interest, because it would not bear what I shall call (...) 'the self-interest relation' to my actual life." (p. 56, my italics)

78 Harman (2009, 2015) makes a similar proposal: "A person may reasonably be glad to have become the person she is; the fact that she does not identify with the person she would have been in the alternative may be sufficient to make her glad to have her actual life rather than the alternative; but this may be so even if the alternative would have been better for her." (2015, p. 324) However, neither Adams

This idea explains why certain relationships, projects, and particular valuable things warrant conservative attitudes about the past: because they have become part of who we are. For instance, the fact that I probably would not have been doing philosophy had my parents emigrated long ago supports my attitude of affirmation in FRANCE, while the fact that in that world I would have bought my shoes in different stores does not, because my history of doing philosophy, unlike my shoe-buying record, is part of my biographical identity. Thus, this proposal does not so much *compete* with personal value or particular value theories, as it offers a deeper unifying explanation of the normative weight of certain personal attachments and particular valuable things in our retrospective outlook.

Importantly, when judgments about our biographical identity support affirmation and lack of regret, there need be no intermediate step where we assign *value* to the relevant parts of our identity. This is, I believe, how we can make sense of disvaluable things in our past as objects of affirmation. Experiences of adversity and hardship, for instance, can become nodal points of the narratives of our lives just as much as the good things in our past. And this can give us reason to affirm them retrospectively, without requiring us to think of those experiences as carrying any new type of value.

This proposal also explains why reasons for regret tend to weaken over time: events that do not define us at a given time may become embedded into our life narrative over the years, and thus may turn into suitable grounds for affirmation later on.

Now we can also explain the potential asymmetry between our retrospective and prospective attitudes: while a commitment to who we are may influence our prospective attitudes—we may be moved to make choices that would fit our self-conception—our future biographical identity is still open, so we cannot be attached to it in the way that we are attached to our past.

What are the criteria for assigning something to our biographical identity, and how much weight should our identity thus understood carry in our retrospective outlook? First of all, we should be wary of any attempt to offer sharp answers to these questions. Our judgments about who we are in a biographical sense are typically imprecise and shifty, and so are the retrospective attitudes grounded in these judgments. Any good theory of these matters should reflect the vast room for indeterminacy and reasonable disagreement that we find in our lived experience.[79]

---

nor Harman says much about what this ethically loaded notion of identity amounts to. One of my goals in this paper is to connect their proposals with recent work on narrative identity.

79 For instance, even though the attitude of affirmation is intelligible and reasonable in a case like FRANCE, regret also seems rationally permissible, depending on how the subject construes his biographical identity and its normative import. We can make sense of our conservative retrospective attitudes by appealing to the central role played by our narrative identities without finding any fault with less conservative sensibilities. This is one important respect in which my view differs from Adams': for him, bitterness about the life one might have had is *groundless* if one's actual life is good and the relevant alternative is thoroughly different—in other words, his verdict is that lack of regret is not only permissible, but *required* in a case like FRANCE. I believe this is too strong, and prefer a permissivist approach on which it can be rational to give more weight to the additional value one finds in an alternative life path, even when weighing it against something that is part of one's biographical identity. Indeed, one might even be *proud* of who one is and still reasonably regret a better life in which his biographical identity would have been different.

That being said, I do not have a substantive theory of biographical identity to offer in this paper. My goal has been to isolate this dimension of our retrospective outlook and some key features of how it interacts with metaphysical and evaluative judgments. However, my proposal dovetails with some recent conceptions of personal identity, like Marya Schechtman's (1996) or David DeGrazia's (2005), who distinguish between metaphysical and ethical questions about identity, and defend a narrative theory of personal identity in the ethical sense. The core idea of this theory, which Schechtman calls the *narrative self-constitution* view, is that we construct our identity by constructing stories of our lives, in which we assign a central role to certain events, experiences, and so on. I do not have the space here to discuss the merits of this theory, or the challenges that it faces,[80] but I see it as a plausible account of the kind of identity that explains and justifies our conservative retrospective attitudes.

Even if the right theory of biographical identity is yet to be properly articulated, the phenomenon I've been talking about supports Schechtman's and DeGrazia's pluralist approach to personal identity, and in particular the separation between metaphysical and ethical questions about identity. Retrospective judgments about who we are arise against the background of settled facts about metaphysical identity. Moreover, these judgments are too open and fluid to obey the logic of numerical identity, and can be weighed against judgments of value, in a way that would be unsuitable for purely metaphysical verdicts. Our conservative attitudes about the past provide thus a new fertile ground for developing and testing non-metaphysical accounts of our identity, centered on the narratives of our lives.[81]

References

Adams, Richard Merrihew. 1979. "Existence, Self-Interest, and the Problem of Evil." Noûs 13:53–65.

Cohen, G.A. 2012. "Chapter 8. Rescuing Conservatism: A Defense of Existing Value." In G.A. Cohen (ed.), *Finding Oneself in the Other*, 143–174. Princeton University Press.

DeGrazia, David. 2005. *Human Identity and Bioethics*. Cambridge University Press.

Kolodny, Niko. 2010. "Which Relationships Justify Partiality? General Considerations

80 For instance, how to resolve the tension between the idea of narrative identity as our own creation, and the sense that the narratives we tell ourselves are meant to capture facts about who we are, and can be off-track as descriptions of those facts. This tension is one reason why I prefer to use the term *biographical identity*. Another reason is that this term does not carry the connotation that our identities should have any particular narrative structure.

81 I am grateful to Sherry Kao, Tom Nagel, and Sharon Street for their helpful comments on previous versions of this paper. I have also greatly benefited from discussions with Harjit Bhogal, Paul Horwich, David Velleman, Dan Waxman, Martín Abreu Zavaleta, and Mike Zhao. Many thanks as well to audiences at New York University, Rutgers University-Newark, the 2015 Conference on Analytical Existentialism at Boğaziçi University in Istanbul, and the 2016 Rocky Mountain Ethics Congress at the University of Colorado Boulder for their useful feedback.

and Problem Cases." In Brian Feltham and John Cottingham (eds.), *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*. Oxford University Press.

Harman, Elizabeth. 2009. "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires."

Philosophical Perspectives 23:177-199.

—. 2015. "Transformative Experiences and Reliance on Moral Testimony." *Res Philosophica* (2):323-339.

MacIntyre, Alasdair. 1984. *After Virtue: A Study in Moral Theory*. University of Notre Dame Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Schechtman, Marya. 1996. *The Constitution of Selves*. Cornell University Press.

Scheffler, Samuel. 1997. "Relationships and Responsibilities." *Philosophy and Public Affairs* 26:189– 209.

Setiya, Kieran. 2016. "Retrospection." *Philosophers' Imprint* 16.

Velleman, J. David. 1996. "Self to Self." *Philosophical Review* 105:39–76.

—. 2015. "Persons in Prospect." In *Beyond Price*, 79–139. Open Book Publishers.

Wallace, R. Jay. 2013. *The View From Here: On Affirmation, Attachment, and the Limits of Regret*. Oxford University Press.

Intelligibility and the Guise of the Good

Paul Boswell

GRIN/CRÉ, Université de Montréal

Abstract:

The Guise of the Good holds an agent only does for a reason what she sees as good in some way. There are two main versions of the theory. According to the attitudinal version, desire has a presenting-as-good character but the good need not figure in desire's content. According to the rival assertoric version, desires are better understood as representations with a normative content that is presented with assertoric force — that is, the force shared by perception and belief. In this paper I present a dilemma for the attitudinal theorist who relies upon one of the main motivations for the Guise of the Good: its ability to account for the intelligi- bility of action for a reason. I show that the very property Guise of the Good the- ories need to answer an objection from Kieran Setiya and Michael Stocker forces them to characterize their view in a way that either favors the assertoric model or cannot explain the intelligibility of action. The upshot is that Guise of the Good theorists should move to assertoric formulations of the view.

Dave claims he had never before considered suicide, and he doesn't recall being de- pressed or even sad at the time. But late one otherwise-ordinary night, while walking down a quiet street, Dave saw a car approaching, and it occurred to him that what he should do is kneel down in the street and be hit by the car. As he later put it, "It seemed to make perfect sense to me. And so that's what I did".1 The car screeched to a halt just in front of him, and a man got out and demanded to know what he was doing. According to Dave, "I looked at him, and I said, I don't know. I had no idea. I had no explanation for him. … I still today — what, 25 years later — don't understand what that was and why I did that".2

In spite of the fact that it seemed in some way to make sense to him to kneel down, in another way it must have been utterly unintelligible to Dave to kneel. After all, he could find no reason for his action. Cases like Dave's thus motivate what is often called the Intelligibility Constraint (IC) on action for a reason:

(IC) If an agent φs for a reason, then φing is intelligible to her (according to a certain sense of "intelligible").

IC is not exactly a platitude, concerning as it does an intuitive but yet-unarticulated no- tion of intelligibility (about which more later). Moreover, it depends upon a particular notion of action for a reason. If, in contrast, after months of psychotherapy Dave were to unearth a repressed fear of appearing to be a failure to his father, and discovered that his genuflection so many years ago was a kind of prayer for forgiveness, then there would be another sense in which he knelt for a reason: he did it in the hope that he would be forgiven. But so long as we stipulate that at the time of action this motiva- tion was fully repressed and that Dave then had absolutely no idea why he was kneeling, then there also a sense in which any such motivation does not speak to a reason of his for kneeling. IC, then, claims that there is a deep connection between this latter sense of acting for a reason and a certain way in which actions may or may not be intelligible to their agents.

One traditional and perennially popular theory in the philosophy of action, the Guise of the Good (GG), is often held by its proponents to explain IC.3 According to GG, an action done for a reason must be seen as good by its agent. Furthermore, it holds that action for a reason must be intelligible to its agent because for an action to be intelligible in the relevant way just is for it to be seen as good in some way.4 Now, it so happens that current GG theories can be divided almost neatly into two camps according to how they think of appearances of the good.5 Some, attitudinal views, hold that these appear- ances have a presenting-as-good character but that normative notions need not figure in the content of these states. Others, assertoric views, hold that these appearances are better understood as representations with a normative content that is presented with assertoric force — that is, the force shared by perception and belief.

In this paper I present a dilemma for the attitudinal theorist who aims to explain IC. In particular I show that the very property that GG theories need to answer an objection from Kieran Setiya and Michael Stocker forces them to characterize their view in a way that either favors the assertoric model or fails to capture the intelligibility of action. The upshot is that GG theorists should move to assertoric formulations of the view.

1 The Guise of the Good

GG is really a family of views which all hold that human action or motivation to act, of some special kind or another, is only possible insofar as the agent acts or is motivated to act because of the good she sees in so acting.6 According to these views, goodness, or apparent goodness, plays a primary role in the motivation and explanation of action.7

Often the view is interpreted as about desire: any action for a reason is motivated by a desire which presents that action as good. For simplicity's sake, that's the version I discuss here.

It's often held among GG theorists that the good is the formal end of desire (e.g. de Sousa 1974, p. 538; Lawrence 1995, p. 130; Tenenbaum 2007, p. 6), where this typi- cally

means that it is an end of desire defined in terms of desire itself. If the good is the formal end of desire then the good may be the desirable, or perhaps simply whatever any desire aims at. However, one must bear in mind that GG requires more than the exis- tence of such a formal end; it must entail that the agent is guided by her mental access to the good, for those who reject GG can accept that there are formal standards for desire (Schroeder 2008).8

The attitude model

The attitude model is a recently popular way of understanding the sense in which desire presents action as good. The point of departure for the attitude model is a supposed analogy between belief and truth on the one hand, and desire and the good on the other. The formal object of belief is truth, it is held, and furthermore to believe is necessarily to regard the propositional content of one's belief as true. But truth cannot plausibly be held to necessarily be part of the propositional content of belief, as if regarding one's belief as true were like regarding a project to be a success, since that would set off a problematic regress. So, truth must figure in the force of belief rather than its content, where 'force' here refers to the representational properties of an attitude apart from its content.9 Truth may be held to figure in the intentional mode of the attitude of belief, that is, in the manner in which belief represents its propositional content: to believe that P is to represent-as-true that P .

According to the attitude model of GG, desire has the same relationship, at least in broad outline, with respect to the good. When one desires that P , one does not desire that it be good that P , for goodness is not part of the content of desire. Rather, desire is held to be a primitive presenting-as-good attitude (Stampe 1987; Tenenbaum 2008;

The assertoric model

On the assertoric model of GG, a desire entails the existence of a state in which evalua- tive content is presented with assertoric force — that is, the force that defeasibly licenses belief or inference.10 Two cognitive, and in a broad sense of the term, representational states can share content and yet differ with respect to force. One's merely imagining that John Rawls is standing in the doorway would not even defeasibly rationalize the belief that John Rawls is standing in the doorway. However a perceptual experience with the same content would. Thus perception and belief, but not imagination, share assertoric force because they both defeasibly license belief or inference. The thought is then that the motivations at the center of GG are to be understood in terms of beliefs or perceptions concerning good actions.

In the assertoric camp belong Raz (1999), Buss (1999), and Gregory (2013), who all hold that action for a reason requires a belief about the good or about one's normative reasons; Davidson (1978, p. 86), who holds that desires express evaluative judgments; and Oddie (2005) and Hawkins (2008), who take desires to be perception-like experi- ences of the good.

2 Intelligibility

Though many GG theorists argue from an alleged intelligibility constraint on action for a reason to the conclusion that actions an agent does for a reason must be seen as good, the argument is rarely set out and defended at length. In this section I present a brief tour of the notion of intelligibility with the aim of thereby rendering IC at least prima facie plausible.

The notion of intelligibility at play is specific in that an action is intelligible to its agent only if the agent has some consideration in mind in virtue of which it is intelligible to him. Take for instance Quinn's infamous Radio Man, who is in a bizarre functional state that causes him to turn on any radio at hand — though he does not turn them on in order to hear anything, or indeed in order that anything else happen. It is just something that he is disposed to do (Quinn 1993, p. 32). What's most conspicuously lacking in Radio Man is an idea of what is to be gained by his turning on radios.

To be sure, an intelligibility requirement built solely on this idea, that in order to

φ for a reason there must be some consideration the agent takes to be the case and in light of which φing is intelligible to them, is neither controversial nor interesting. Quite plausibly it falls out of the very concept of acting for a reason that one have at least some reason in mind in acting. But the special interest in the GG theorist's conception of intelligibility is that not just any feature, consequence, or aspect thought by the agent to be instantiated by the action suffices to make that action intelligible to the agent:

> Suppose, for example, that you notice me spray painting my shoe. You ask why I am doing that, and I reply that this way my left shoe will weigh a little more than my right. You ask why I want the left shoe to weigh a little more. Now suppose I just look at you blankly and say, "That's it." I seem not to understand your puzzlement. You grasp for straws. "Is this some sort of performance art, on the theme of asymmetry?" "No." "Is someone going to weigh your shoes as part of some game?" "No. Why do you ask?"11

Here it is clear that there is an oddness about the interlocutor's explanation even though they have correctly identified a plausible consequence of spray painting their shoe.

So a crucial aspect of this conception of intelligible action is that not just anything the agent could, conceivably, sincerely say about her reasons or motives is eligible to articulate the point she sees in acting. Some properties, when taken by their agents to be instantiated by an action, can make that action intelligible, and others cannot. Being instrumental to something one is compelled to do does not make an action intelligible. Nor does thinking of one's action as possessing a thin normative property, such as being good to do as such, render it intelligible, as Dave's example shows. So if indeed action for a reason must be intelligible to its agent, what explains why such action can occur under the guise of its instantiating certain properties and not others? GG proposes just such an explanation: the property must be, as Anscombe put it, "one of the many forms of good" (op. cit., p. 77), i.e. a particular or substantive good.

Once we clarify that value can render an action intelligible only when an agent has a particular good in mind, we can neatly sidestep an otherwise worrisome objection raised by both Michael Stocker (2004) and Kieran Setiya to GG's explanation of IC. The objection centers around the fact that merely noting that someone's action was performed under the guise of the good does not suffice to make the action intelligible. To paraphrase a dialogue in Setiya (2010, p. 97):

"She is drinking coffee because she loves Sophocles." "What? That makes no sense at all.

"Oh yes it does. She thinks that makes it good to drink coffee."

What has gone wrong here is that we are not given any substantive good, no particular value, under whose guise the coffee-drinker drinks. If we were told that she's drinking coffee because she thinks this honors Sophocles, we may be confused as to why she'd think that, but we would then appreciate her goal of honoring a literary figure.12

3 Against the attitude model

According to the attitude model it is constitutive of desire that it presents-as-good the desired action, in much the same way that belief presents-as-true its content. Generally speaking, proponents of this version hold that the good so presented is the formal object of desire.

But it would appear that this reliance on a formal notion of the good is also what prevents the attitude model from accounting for the intelligibility of action. As we saw above, if the guise of the good is to explain the intelligibility of action, it must explain it by reference to the appearance to the agent of particular goods. That was precisely what we needed to respond to Setiya and Stocker's objection above. But all that the attitude model can secure is that when an agent desires an action, it appears to them to be formally good, that is, to meet the formal aim of desire. And it seems that a formal aim of desire as such would need to be non-substantive and characterized in terms of a thin normative notion. On this view, the desired action appears (e.g.) choiceworthy, desirable, or the thing to do as such. The attitude model thus secures the intelligibility of desired action only if Dave's action is intelligible to him. For after all, it appeared to Dave that what he should do is kneel down and be hit by the car. Clearly it would have made his action no more intelligible if he had thought, say, The thing to do is to kneel down and be hit by the car. But Dave's action wasn't intelligible to him, so the attitude model cannot be used to secure the intelligibility of desired action.

One GG theorist appears at first to have the resources to deal with this problem.

According to Sergio Tenenbaum, desires not only present their contents as meeting the formal aim of (any) desire, but they present it under a certain perspective (Tenenbaum 2007, §1.5). According to Tenenbaum, the perspective under which one desires explains the particular way in which one's desiring is intelligible. The bare fact that Sue wants to damage Ms. S's boat does not intelligibly explain why Sue is throwing stones at it.

But if we are told that Sue wants to damage Ms. S's boat out of envy, it appears we have been given an explanation that shows her stone-casting to be intelligible.13 Other perspectives that can make certain actions intelligible include honesty, being a cinephile, a gourmand, or a good parent — all plausibly construable as organized around specific goods.

We can also work out a more general version of the strategy at work here. The attitudinal strategist may weaken the proposed analogy with truth in a certain respect. According to the story told in §1, the connection between belief and truth is constitutive of and common to any belief, for all beliefs present-as-true their content. But perhaps token desires may present different goods.14 Perhaps it is constitutive of desire that it present some good or other, even though no particular good may be required. Truth is one and good many, it might be said. Some such necessary connection between desires and specific goods is needed to overcome the Setiya/Stocker objection, and it also seems to be what Tenenbaum is trying to secure.

Unfortunately, such a connection is not ultimately available to the attitudinal the- orist. I'll consider two versions of this view. The first characterizes the connection between the attitude of desire and a specific good in terms of an adjective embedded within a verb: appetite presents-as-tasty a treat. The second, Tenenbaum's way, charac- terizes it adverbially: Sue desires enviously. Against the first I'll argue there is no obvious, good reason to think that it characterizes an aspect of an attitude like desire as opposed to its content, and that there is reason to think the opposite. Against the second, I argue that such a connection would not secure the needed mental access to the specific good which GG requires.

Content and attitude

What is an attitude? Orthodoxy has it that it is a relation between a subject and a content, particularly a propositional content.15 Most of the disagreement on this question is over the deeper nature of this relation. Staying at the level of commonly accepted platitudes, we could say that the content of a mental state gives what the mental state presents to the mind, what it is about, and the attitude provides how that content is presented, or the way in which the subject takes that content or that presentation.

So when the attitudinal theorist tells us the desire for a treat presents-as-tasty the treat, should we take her at her word that she has characterized an aspect of the attitude of desire? Against this, note that nearly all the substantive, intelligible-making goods, such as health, success, beauty, and tastiness, are properties of the things, states of affairs, and people that our motivations are concerned about. If it is true that your appetite for the treat presents-as-tasty the treat, then it seems we can also express your attitude by saying

that you take it as true that the treat is tasty. You in some way attribute tastiness to the treat; that's why it's intelligible to you to eat the treat. And of course, we would naturally think that 'tasty' figures in specifying the content of this attitude. But it is distinctive of attitudinal views that the good is not supposed to play a role in motivation by figuring in the content of a mental state.

But if we do not understand this adjectival locution in terms of attributing a value property to things that might bear it, it is hard to make sense of it. What else could it mean to take-as-beautiful a painting? What sort of manner of response to a content is that? Perhaps it means whatever manner is appropriate to a beautiful object.16 But why would that manner necessarily imply that the subject has mental access to the relevant good, as GG requires?

The adverbial version of the attitudinal view shares a similar epistemic difficulty, so it is to that I now turn.

Mental access

On Tenenbaum's view, Sue's desire to throw rocks at Mrs. S's boat has an adverbial characterization: she desires enviously to throw rocks at it. Judging by Tenenbaum's development of the view, it is clear that this is a refinement of the attitude strategy. In order to act out of envy Sue need not have the explicit aim of acting out of envy; envy need not enter into the content of her mental states. She does not necessarily desire to be envious. Rather the envy is held to figure in the manner in which the world and practical possibilities appear to her. She enviously desires the destruction of the boat, and this manner of appearance is made manifest in the options for action she takes seriously, her irritability towards praise of the boat, the comments she makes about Ms. S, etc.

However, it turns out that this maneuver does less good for the attitude theorist than she might have hoped. Clearly there are such modes of acting and desiring, and we can appeal to them in order to give third-personal explanations of an agent's action by clarifying their reasons for it and making it intelligible to us as spectators. But to say that an agent acted from a certain perspective is not always to explain the action in terms of the agent's point for that action. Sue might be self-consciously envious, but it's also possible that she is completely ignorant of her own enviousness. Her conscious motivations may have been limited to thoughts that Ms. S was fundamentally at fault, and that she needed to be taken down a peg. One can easily imagine a friend confronting Sue about her enviousness and Sue, upon realizing the truth about herself, coming to terms for the first time with her own envy. This reckoning would be rather like Dave's discovery in psychotherapy that he was motivated to lie down in the street by his repressed fear of failure: both agents may be said to have learned about their reasons for acting, but not in the sense of 'reasons for acting' we are looking for, since in both cases the motivation (envy or fear) was opaque to the agent.17 Thus, while it may be that when one acts for a reason, one acts under a perspective on the good, this latter cannot by itself explain the intelligibility of action for a reason.

Nor is this problem unique to Tenenbaum's particular strategy. In this short version of the paper, I will briefly outline what I take to be the main problem with this view:

1. In acting for a reason one is guided by a desire that gives one mental access to the (apparent) goodness of so acting.

2. When an agent φs for a reason, φing is intelligible to her in virtue of a state of mental access to the (apparent) good of φing.

3. In cases relevant to GG, mental access to implies consciousness of.18

4. Consciousness of anything is consciousness of a content.

5. So, goodness figures in the content of the state that makes action for a reason intelligible.

As noted above, premise 1 states an implication of GG. Premise 2 captures how GG explains IC. Dave's case is extremely good evidence that without conscious access to a specific good, an agent does not have the right kind of mental access to the (apparent) good of his action, justifying premise 3.19 Premise 4 is quite generally taken as a starting- point in the philosophy of mind: the contents of consciousness just are those things of which we are conscious. From these premises, the conclusion in 5 follows. But 5 contradicts attitudinal views that aim also to account for the intelligibility of acting, according to which goodness need not figure in the content of desire.

4 Conclusion

IC provides a prima facie attractive motivation for the Guise of the Good. After all, it does seem to be the case that actions must appear to their agents to have specific properties in order for them to be intelligible, and that we can explain which properties do make actions intelligible by reference to specific goods. But it turns out that one of the major classes of GG theories, attitude theories, cannot explain the intelligibility of action for a reason, contrary to the claims of many of its proponents. In order to reply to the Stocker/Setiya objection, these theorists need to appeal to a specific good that makes action intelligible. But this leaves them with a dilemma: either they characterize the connection to the specific good in a way that makes an assertoric view more plausible, or they fail to explain the agent's mental access to the good, as GG requires. This finding provides support for non-attitudinal views of GG.

Notes:

1 Recorded in Glass (2002).

2 Ibid.

3 See Anscombe (1963); Quinn (1993); Raz (1999, 2010); Tenenbaum (2007); Sussman (2009); Boyle & Lavin (2010, pp. 188-9), among others.

4 See Anscombe (1963, pp. 70-8); Tenenbaum (2007, pp. 32-3).

5 Schroeder (2008) and Schafer (2013) refer to similar distinctions.

6 See Tenenbaum (2013) and Orsi (2015) for recent overviews. Note that although "sees that P " often has a factive sense in English, here that sense is not presumed.

7 See for instance Raz (2009) on what he calls the "normative/explanatory nexus".

8 The 'good' in 'Guise of the Good' should in fact be seen as a technical notion, since the family comprises Guise of Reasons theorists and Guise of Ought theorists as well. For more on the notion of normativity at play in GG theories, see Boswell (2016), Ch. 3, §2.

9 Ultimately the notion of force derives from Frege. See for instance Frege (1918, p. 294).

Schafer 2013; Kriegel 2017), or goodness is considered part of the form of desire (Saemi

2015), or it is suggested that it is constitutive of a desire that P that it aims to get it right as to whether P is good (Velleman 1992, though Velleman goes on to reject GG).

10 I take the notion of assertoric force from Schafer (2013).

11 Clark (2010, pp. 234-5).

12 One might think that Setiya's objection fails for a different reason, that quite generally possessing the testimony that P need not make P intelligible to one. But quite plausibly, this general truth relies on a distinct kind of intelligibility. One can hear testimony that P without understanding what is said, and so find the utterance unintelligible in that sense. But in Setiya's dialogue we are to imagine that the interlocutor does understand the last line and yet finds the agent's action unintelligible nonetheless.

13 Ibid., p. 43; Tenenbaum's example.

14 Here I use "presentation" and its cognates in Brentano's sense as a merely contentful state that of itself entails no commitment to the truth, appropriateness, etc. of its content, and which can can be taken up into an attitude that does entail some such commitment. See Brentano (1874, p. 61ff.).

15 E.g. Fodor (1978), generalizing slightly.

16 As noted above, many attitudinal theorists take the presenting-as-good character of desire to be primi- tive. The response just considered is unavailable to these theorists, who are unable to explain this character in more basic terms and thus face extraordinary difficulty explaining why 'good' should be thought to char- acterize an aspect of an attitude apart from its content.

17 Indeed, Tenenbaum seems to acknowledge that the perspective under which one acts can sometimes be opaque to the agent; see ibid., p. 50.

18 Consciousness of x does not here imply the existence of x. The sense is that in which I am conscious

of a red table in front of me even when I hallucinate it.

19 Note the contrast with belief here: it is not at all clear that belief involves mental access to the truth of a proposition over and above the proposition itself, and it is even less plausible that belief that P requires consciousness of the truth of P . There is a sense in which one is generally guided by defeasibly justified inferences when believing, but this guidance need not take the form of an appearance of truth.

References

Anscombe, G. E. M. 1963. Intention. Blackwell, Oxford, UK, second edn.

Boswell, Paul. 2016. Affect, Representation, and the Standards of Practical Reason. Dissertation, University of Michigan, Ann Arbor, Ann Arbor, MI.

Boyle, Matthew & Douglas Lavin. 2010. "Goodness and Desire." In Desire, Practical Reason, and the Good, Sergio Tenenbaum, editor, 161–201. Oxford UP, New York.

Brentano, Franz. 1874. Psychologie vom empirischen Standpunkte. Duncker & Humblot, Leipzig.

Reprinted in English translation in: Brentano, Franz. 1995. Psychology from an Empirical Standpoint. Ed. Oskar Kraus and Linda L. McAlister, trans. Antos C.Rancurello, D.B.Terrell, and Linda L. McAlister. Routledge, London, UK.

Buss, Sarah. 1999. "What practical reasoning must be if we act for our own reasons." Australasian

Journal of Philosophy, vol. 77 (4): 399–421.

Clark, Philip. 2010. "Aspects, Guises, Species, and Knowing Something to Be Good." In Desire, Practical Reason, and the Good, Sergio Tenenbaum, editor, 234–244. Oxford UP, Oxford.

Davidson, Donald. 1978. "Intending." In The Philosophy of History and Action, Yirmiaku Yovel, ed- itor. D. Reidel and the Magnes Press, Jerusalem, second edn. Reprinted in Davidson, Donald.

2001. Essays on Actions and Events, 2nd edition. Oxford UP, Oxford: pp. 83-102. References to this edition.

Fodor, Jerry A. 1978. “Propositional Attitudes.” The Monist, vol. 61 (October): 501–23.
Frege, Gottlob. 1918. “The Thought: A Logical Inquiry.” Mind, vol. 65 (1): 289–311.

Glass, Ira. 2002. “Devil on My Shoulder.” Episode 213, URL http://m.thisamericanlife.org/radio-archives/episode/213/transcript.

Gregory, Alex. 2013. “The Guise of Reasons.” American Philosophical Quarterly, vol. 50 (1): 63–72. Hawkins, Jennifer. 2008. “Desiring the bad Under the Guise of the Good.” The Philosophical

Quarterly, vol. 58 (231): 244–264.

Kriegel, Uriah. 2017. “Will and Emotion.” In Mind and Reality in Brentano’s Philosophical System,

–. Oxford University Press. Draft; forthcoming 2017.

Lawrence, Gavin. 1995. “The Rationality of Morality.” In Virtues and Reasons: Philippa Foot and Moral Theory, Rosalind Hursthouse, Gavin Lawrence & Warren Quinn, editors, 89–148. Oxford UP, Oxford.

Oddie, Graham. 2005. Value, Reality, and Desire. Oxford UP, Oxford, UK.

Orsi, Francesco. 2015. “The Guise of the Good.” Philosophy Compass, vol. 10 (10): 714–724. Quinn, Warren S. 1993. “Putting rationality in its place.” In Value, Welfare, and Morality, Christo-

pher W. Morris & R.G. Frey, editors, 26–49. Cambridge UP, Cambridge.

Raz, Joseph. 1999. “Agency, Reason, and the Good.” In Engaging Reason, 22–45. Oxford UP, New York, NY.

—. 2009. “Reasons: Explanatory and Normative.” In New Essays on the Explanation of Action, Constantine Sandis, editor, 184–202. Palgrave Macmillan, New York.

—. 2010. “On the Guise of the Good.” In Desire, Practical Reason, and the Good, Sergio Tenenbaum, editor, 111–137. Oxford UP, New York, NY.

Saemi, Amir. 2015. “Aiming at the good.” Canadian Journal of Philosophy, vol. 45 (2): 197–219. Schafer, Karl. 2013. “Perception and the Rational Force of Desire.” The Journal of Philosophy,

vol. 110 (5): 258–281.

Schroeder, Mark. 2008. "How Does the Good Appear To Us?" Social Theory and Practice, vol. 34 (1): 119–130.

Setiya, Kieran. 2010. "Sympathy for the Devil." In Desire, Practical Reason, and the Good, Sergio

Tenenbaum, editor, 82–110. Oxford UP, New York, NY.

de Sousa, Ronald. 1974. "The Good and the True." Mind, vol. 83 (332): 534–551.

Stampe, Dennis W. 1987. "The Authority of Desire." The Philosophical Review, vol. 96 (3): 335–

381.

Stocker, Michael. 2004. "Raz on the Intelligibility of Bad Acts." In Reason and Value: Themes from the Moral Philosophy of Joseph Raz, R. Jay Wallace, Philip Pettit, Samuel Scheffler & Michael Smith, editors, 303–332. Oxford UP, New York.

Sussman, David. 2009. "For Badness' Sake." Journal of Philosophy, vol. 106 (11): 613–628. Tenenbaum, Sergio. 2007. Appearances of the Good. Cambridge UP, Cambridge, UK.

—. 2008. "Appearing Good: A Reply to Schroeder." Social Theory and Practice, vol. 34 (1): 131–

138.

—. 2013. "The Guise of the Good." In The International Encyclopedia of Ethics, Hugh LaFollette, editor. Wiley, Hoboken. URL http://philpapers.org/archive/TENGOT.pdf.

Velleman, J. David. 1992. "The Guise of the Good." Noûs, vol. 26 (1): 3–26. Reprinted in Velleman, J. David. 2000. The Possibility of Practical Reason. Oxford UP, New York, NY. References to this edition.

mbaron@indiana.edu

SEXUAL CONSENT, REASONABLE MISTAKES, AND THE CASE OF ANNA STUBBLEFIELD

Marcia Baron
Indiana University

1. INTRODUCTION

For the crime of rape, a false belief that the other party was consenting should provide a complete defense[82] only if the belief is reasonable. This is a common view, one I defended years ago and hold without hesitation. I won't be defending it in this paper.

Just what should count as a reasonable belief is a further question, and in general the tendency is one of excessive generosity. At least that is my experience in reading cases: I almost never read a court ruling in a sexual assault case where I think the mistaken belief that was deemed unreasonable perhaps should have been considered reasonable, and I fairly often come across a case where the belief, apparently deemed by jurors or judges to be reasonable,[83] strikes me as clearly unreasonable. A recent conviction is one of those rare cases where there are grounds for thinking that the belief should have been considered reasonable. Not that I'm *sure* it should be deemed reasonable; and my concern is not primarily to argue that it should, but to reflect--and generate discussion--on what sorts of considerations should weigh in. The case provides food for thought on the question of what should count as a reasonable (but false) belief that the other party was consenting.

Before turning to the Stubblefield case, I should explain my starting points on rape law--on what I think the law should be.

The first I already noted: I think that a false belief that the complainant was consenting should be a complete defense only if it was reasonable. I also think there should be no force requirement. The actus reus (act element) of rape should be understood to be *nonconsensual sex*. {Not forced nonconsensual sex; not even forced sex (where non-consent does not have to be proven, only force does)[84]. It should be

[82] I put it this way because to put it accurately would involve more legalese than is desirable, but I should note that technically it doesn't count as a true defense; rather, the idea is that the mens rea requirement has not been met. The elements of the offense have not been proven beyond a reasonable doubt. See Dressler, *Understanding Criminal Law*, on the difference. [Complete reference later.]

[83] Or at least, they judged that there is room for reasonable doubt as to whether it was unreasonable.

[84] This is, however, a much more plausible option than requiring both force and consent, and more plausible still is to understand rape as coerced sex, following Scott Anderson (who distinguishes 'coerced' from 'forced' and argues that understanding rape as coerced sex is helpful for explaining the "distinctively gendered pattern of incidence" of rape, "its distinctively devastating harms to women" (80) and "why the sexual aspect of these crimes is especially problematic for victims" (81) ["Conceptualizing Rape as

nonconsensual sex.} Finally sexual consent should be understood as distinct from wanting or desiring sex. This is a less common point than the other two, so some elaboration and defense are in order.

2. CONSENT (OR: WHY NONCONSENSUAL SEX AND UNWANTED SEX ARE NOT IDENTICAL)[85]

Consent--whether to sex or to something else--is something one does (or in its noun form, something one gives).[86] It is not best understood as something one feels, and consenting should not be equated with wanting or desiring. To consent to something is to agree to it (and to agree under conditions where one is reasonably free to decline). *A* can desire something without agreeing to it—as, indeed, with sex, where *A* feels a strong desire for sexual intimacy with *B* but for such reasons as that A is married to *C*, *A* declines *B*'s invitation. One can also agree to something without desiring it (as when one agrees to give a housemate a ride to the Metro station, realizing that the other's need for the ride is greater than one's own need to continue, without interruption, the translation one is working on).[87]

This is so far just a conceptual point about consent, and one might say: What's in a word? Even if consent and desire are not the same thing, might it not be useful for purposes of the law to understand sexual consent as sexual desire? Point well taken, but I don't see that it *is* useful. On pragmatic grounds as well it is better to differentiate sexual consent from sexual desire and to understand sexual consent as first and foremost something one gives, not something one feels. Wishful thinking combined with arrogance can easily support the thought, "I know she said 'no' but I can tell she really wants it and I guess she doesn't really know her own mind, or maybe she just has a hard time saying what she wants." If sexual consent and sexual desire are equated, the initiator may be well situated to say, "Yes, she consented! She really did want it, though she denied it." This is true whether we understand the desire in question to be sexual desire or an all things considered desire to have sex (on this occasion and with this person), though the risk is probably greater if sexual consent is equated with sexual

---

Coerced Sex," *Ethics* 127 (2016): 50-87]). It is beyond the scope of my paper to work through the comparative advantages and disadvantages of understanding rape as nonconsensual sex and understanding it as coerced sex, and as Anderson observes, even if his account is accepted, the question of consent, and in particular, mistakes concerning consent, remains. Specifically, on his account, it would be a defense to argue that the defendant "had reason to believe that" the complainant "consented to his activities" (83).

[85] There is, to be sure, far more to say on the topic and a considerable literature that I do not engage with here. For a far more thorough defense of an understanding of consent as requiring communication, see Tom Dougherty, "*Yes* Means *Yes*: Consent as Communication," *Philosophy and Public Affairs* 43 (2015): 224-253.

[86] We can classify it as a performative, as long as we don't take it to be a requirement of a performative that there is proper uptake. The idea should *not* be that my non-consent counts as such only if it is so understood by the person to whom I convey it!

[87] One might contest this, claiming that if one agrees to it, even if one would in some sense rather work on the translation, one must want to provide the ride more than one wants to stay in and work on the translation. Clearly this hangs on just how we understand 'want', an issue I will sidestep. I am less concerned to convince readers of the conceptual point than to bring out the pragmatic reasons in favor of distinguishing consenting from wanting or desiring. That 'want' and 'desire' are so slippery, and that consent if understood as something one does rather than something one feels is less slippery, are strong reasons for understanding consent accordingly.

desire.[88] By firmly distinguishing consent from both sexual desire and an all things considered desire to have sex, we can make it clear that it is unacceptable to override the other person's refusal or lack of consent with one's own judgment of what she (or he) really wants.

If *A* declines *B*'s invitation, that *B* thinks *A* really wants to have sex with *B* is--or should be--neither here nor there. But if desire is treated as equivalent to consent, the option to refuse sex is dangerously undermined. Does *A* have to convince *B* that she doesn't want it, if *B* maintains that she does? One wants to have one's refusal taken seriously, recognized as authoritative.[89] *B* might be in a suitable epistemic position to question whether *A* wants what she says (and thinks) she wants; but that *A* may be wrong about that has no bearing on whether *A* consented. It is something *B* could bring up in an effort to try to persuade *A* to change her mind, but not to show that she *has consented*, or *is consenting.*

It is worth noting that it is not only *A*, the person refusing, who benefits from a distinction between consenting to *X* and desiring *X*. Consider the matter from *B*'s perspective, the perspective of the person initiating sex.[90] *B* is better off if he or she can be confident that *A* really is consenting when she says 'Sure!' As long as the conditions under which *A* has said 'Sure!' are not such as to raise worries that this was not consent but submission out of fear (or feeling so pressured that *A* feels she is being given no choice) and as long as there was no obvious sign of fear or revulsion or something else that suggests that this might well not be consent, *B* should not be called upon--legally required--to check that *A* means what she said and is not merely submitting. There might be unclarities to sort out ('Sure' to which sexual activity?), but there should not be a legal requirement to try to figure out what *A* really wants. Moreover, the fear that later *A* will regret it and believe (possibly correctly, possibly not) that she never did want it, and if upset enough will go to the police {a fear, I gather, of many men regarding initial encounters} should have less of a foothold if it is made clear that what is crucial, for purposes of the law, is that she consented, not that she desired to have sex with *B*. —

{Note: this is not to say that if one notices the other party doesn't appear to want to do what she has said 'Yes' to one shouldn't worry, since after all, she said 'Yes'. It should raise concerns about whether she said 'Yes' only out of fear. Moreover, one might well wish not to have sex with someone who seems unenthusiastic. But that it turns out that she didn't want to does not itself show that she did not consent.}

For purposes of holding people responsible for knowing whether another consented, it makes sense to understand consent as something that, like promising, is "done"—

---

[88] Greater because it may be pretty obvious that the other person is sexually aroused, and inferring sexual desire from arousal is more warranted than is inferring an all things considered desire to have sex. The view that sexual desire constitutes (or at least suffices for) consent is very likely one reason for rape victims' hesitation to report the crime to the police, and it would not be surprising if this were a particularly serious obstacle to male victims reporting an assault (particularly if the victim knows, or believes, that he was visibly aroused).

[89] On the matter of authority, see David Archard, "The Mens Rea of Rape: Reasonableness and Culpable Mistakes," in Keith Burgess-Jackson, ed., *A Most Detestable Crime: New Philosophical Essays on Rape* (OUP, 1999), esp. pp. 222-226.

[90] Not that it is always the case that one person is the initiator, but these cases are of greater interest for my purposes because they are more likely to give rise to mistakes about consent.

something conveyed to another (though not necessarily with proper uptake)[91]—rather than as desire (or any other mental state).[92] Why expect mind-reading? And indeed in other areas of the law where consent converts something that would otherwise be a crime to something either innocuous or positively good, we understand consent as something conveyed, not as something felt. Whether you borrowed my car or are guilty of joy-riding is (as far as the actus reus goes) a matter of whether I consented, and my consent is not a matter of how I feel or felt. {If I told you that you were welcome to use my car on such-and-such a day, I consented to your doing so,[93] even if your request occasioned in me resentment and annoyance. How I felt isn't the issue; it is a matter of whether I told you that you could use it (without it being the case that you forced me to agree to it). If I wanted to let you borrow it, but told you that you couldn't (perhaps because I knew another party would be upset if I lent it to you), I did not consent, despite my warm feelings at the prospect of lending it to you.[94]}

{A further point in favor of distinguishing between unwanted sex and nonconsensual sex is that without the distinction, it is hard to do justice to the importance of not having sex with someone too intoxicated to give consent, or with someone below the age of consent. In each case the person might in fact want to have sex (and with the relevant party). Yet we know better than to think that because a 13-year-old wants to have sex with *B* he or she is consenting to it; ditto for the intoxicated person.}

3. When is a Mistaken Belief that *A* Consented Reasonable?

Obviously, the question does not admit of an answer that will provide anything close to a formula. All we can hope for are some guiding considerations.

To come up with some guiding considerations, let's look at a case where (or so I will argue) a mistaken belief that the other party was consenting should not count as reasonable. I purposely am choosing a case where it is not glaringly obvious that it should count as an unreasonable belief; those where it is are less likely to tell us anything helpful about the cases about which we are not already fairly sure what to say.[95]

---

[91] See note 5, above.

[92] For discussion of some complexities not addressed here, see Peter Westen, *The Logic of Consent: The Diversity and Deceptiveness of Consent as a Defense to Criminal Conduct* (Ashgate, 2004).

[93] Provided that you did not threaten me.

[94] This last paragraph is a lightly revised version of a paragraph in my "Gender Issues in the Criminal Law," in John Deigh and David Dolinko, eds., *The Oxford Handbook for the Philosophy of Criminal Law* (OUP 2011).

[95] However, it is worth noting a certain sort of case, exemplified by both *Morgan* [*DPP v. Morgan* (1975) 2 All ER, pp. 347-383] and *Cogan* [*R v. Cogan* (1975) 2 All ER, pp. 1060-1063]: there the defendant (defendants in the case of *Morgan*) supposedly believed that the complainant was consenting because of what the defendant's husband had claimed. In each case the husband had in effect said to the man/men that whatever she did or said, she really wanted to have sex with him/them (and he also equated wanting to have sex with consenting to have sex). Let's set aside the fact that in *Morgan* the defendants clearly did not really believe this, and imagine that they did. (The jury in *Cogan* did hold that he really believed it--or at least that it wasn't beyond a reasonable doubt that he did.) But they believed it only because (a) of what a third party said, and (b) in accepting the third party's claim, they took it that absolutely nothing she did or said would count as refusing to have sex with them. Clearly it is unreasonable to believe *S* consents if one has antecedently decided that absolutely nothing *S* says or does will count as

The case I have in mind is *R. v Tawera,* a New Zealand case from 1996.[96] Tawera was convicted of raping his 16 year old cousin, who was living with Tawera and his family. (He was 48 at the time.) The appellate court overturned the conviction and directed that a verdict of acquittal be entered.[97] Overturning a conviction for reasons other than procedural grounds is unusual, as the court acknowledges: "[T]his is one of those rare cases when the verdicts cannot be supported, and...a reasonable assessment of the relevant evidence as a whole must have left a tribunal of fact with a reasonable doubt on this essential element."[98] The essential element was that the appellant *not have believed on reasonable grounds that the complainant was consenting*.

I find this an interesting case because the relevant statute is quite sound: There was no force requirement (either as a separate requirement or for non-consent), only a requirement that the sexual connection took place "without the consent of the other person" and "without believing on reasonable grounds that the other person consents to that sexual connection" (s 128). Moreover, s 128A lists "matters that do not constitute consent to sexual connection." Notably, "The fact that a person does not protest or offer physical resistance to sexual connection does not by itself constitute consent to sexual connection for the purposes of section 128 of this Act."

It is an interesting case as well because the jury rendered what seems to me clearly to be an appropriate verdict, and yet the appellate court overturned it on the grounds that it cannot be supported.

What was the evidence, such that it was deemed insufficient to support the verdict? The evidence (not in dispute) was as follows: after the complainant got into her bed, Tawera got into the bed with her, uninvited. He initiated some intimacy; she showed no interest but also did not resist beyond trying to turn her head away when he attempted to put his tongue into her mouth (resistance which he overcame) and trying to hold her thighs together (again resistance he overcame). Her reaction, apart from the resistance just mentioned, was one of passivity, including giving no response at all when he asked "Honey can I stick it in can I stick it in?"

Given that New Zealand law specifies that passivity by itself does not constitute consent (and given that there is no force requirement), it is clear that this was nonconsensual sex. For there was nothing other than a lack of resistance, verbal or physical, to point to as a reason for thinking she consented. (In fact as noted, she did resist; but even if she had been completely passive, that would not have constituted consent.) The court does not claim that it was consensual (nor that the jurors were wrong to conclude that the prosecution had proven beyond a reasonable doubt that it was nonconsensual). But the court denies that the prosecution showed beyond a reasonable doubt that the mens rea requirement was met. Their reasoning emerges when they offer a guess as to how the jurors could have arrived at a guilty verdict:

---

non-consent.

[96] *R v Tawera* 14 CRNZ 290 (1996)

[97] Not that he was entirely a free man; he had also been charged with "having sexual intercourse with a girl under his care and protection." On that charge, no verdict had been taken, and the court ordered a new trial.

[98] *Tawera*, p. 293.

> It may be that the jury became unduly concerned about the direction (correctly given) on s 128A and the fact that a failure to protest or offer physical resistance does not by itself constitute consent. That kind of consideration may of course be highly relevant to whether there was consent, but it does not really bear on the critical issue of belief in consent.[99]

Not on the issue of whether the defendant believed that the complainant was consenting, but surely, I should think, it bears on whether the belief was reasonable!

If the law spells out that *X* does not suffice to constitute consent, and the defendant believes solely because of *X* that she consented, this cannot be a reasonable belief. Perhaps it could if we counted ignorance of the law as an excuse, but we don't (except in rare circumstances, not relevant here.)[100] Treating that as a fixed point, it matters that a mistaken belief that because she did not resist, she consented, is a mistake of law.

The elements of the crime[101] for which he was convicted seem clearly to have been proven beyond a reasonable doubt. The court raises no worries concerning the act element, but holds that the mens rea was not proven beyond a reasonable doubt. However, if 'reasonable' in 'reasonable grounds' is going to play a role in s 128, surely this mistaken belief that she was consenting is *not* based on reasonable grounds. It is based only on the fact that she didn't resist (or more precisely, didn't resist much).

{Now, were there no stipulation in the law that passivity does not by itself constitute consent, there would be some basis for arguing that although Tawera should have realized it was unlikely that she would consent or was consenting (since she is 16 and he 48, she is living in his home, and they are cousins[102]), and should have stopped his advances in the absence of any indication of consent, still, the bar for reasonableness needs to be set low. Morally, sure (one might argue),[103] but we are talking about legal culpability here. For that the bar for reasonableness needs to be lower, and it would be too harsh to say that he acted unreasonably in thinking that more than this was needed (legally) for consent. I am not sure I am on board, but it is not implausible. But since s 128A makes it quite clear that passivity alone does not suffice to constitute consent, surely it cannot be reasonable to substitute his own ideas of what suffices for consent.}

Reflection on *Tawera* suggests two points concerning when a mistaken belief that the other person is consenting is reasonable:

---

[99] *Tawera*, p. 293.

[100] For general information on when mistakes of law can and when they cannot be defenses, as well as an explanation of the distinction between mistakes of fact and mistake of law, see Joshua Dressler, *Understanding Criminal Law*, 7th ed. (Lexis-Nexis, 2015) [add sect. no.]; for a discussion of mistakes of law in connection with sexual consent, see Mayo Moran, *Rethinking the Reasonable Person* (OUP, 2003), pp. 295-299.

[101] He actually was convicted of two crimes, the other being sexual violation by unlawful sexual connection (in this instance, oral sex); I'm simplifying by focusing on just one, since the latter raises no additional mens rea issues.

[102] "She was a daughter of a cousin of the appellant's" (*Tawera*, p. 291).

[103] However, if we are talking about what he should have done morally, not only should he have stopped his advances in the absence of any indication of consent; he should not have initiated any such activity in the first place.

First, the belief had better not be at odds with what the law tells us about what does, or does not, constitute consent. The mistake has to be a mistake about a matter of fact (e.g. how old the other party is, or whether (s)he is only slightly tipsy rather than intoxicated, or what (s)he said), not about a matter of law.

Second, in thinking about whether the defendant's belief was unreasonable, we should attend to any steps (s)he took to ascertain whether the other party was consenting. It is in Tawera's favor that he did ask permission;[104] the problem is just that when she didn't reply, he either chose to take that to be a 'Yes' or decided not to worry about why she was not replying.[105]

That in determining whether the mistaken belief should count as reasonable we should take into account any such steps is part of UK law (and has been ever since the law was revised to require that the mistaken belief be reasonable), and I think it should be part of all sexual assault laws. The 2003 Sexual Offences Act specifies: "Whether a belief is reasonable is to be determined having regard to all the circumstances, including any steps A has taken to ascertain whether B consents."[106]

I turn now to the Stubblefield case, and to the question of whether her belief that D.J. was consenting should perhaps count as reasonable.

4. The Stubblefield Case

Anna Stubblefield was convicted in October 2015 of two counts of aggravated sexual assault of D.J., a disabled man. She claims they had consensual sex. Whether he says so too (as she says he did) depends on whether it is his thoughts that were expressed through the method called "facilitated communication."

There are two bases for the position that this was nonconsensual sex. One is that D.J. isn't capable of consent because he is too impaired cognitively to consent to sex. {It is an interesting question whether there is anything else he could consent to, or whether we should just say he is incapable of consent, period.[107]} The other is that even if he is not

---

[104] Regarding vaginal penetration; there is no indication in the ruling that he also did so regarding oral sex. (See n. 20, above.)

[105] Indeed, compared to another defendant from two years earlier who climbed on top of a sleeping woman, quickly pulled down her pants and penetrated her, Tawera looks pretty good! At least Tawera gave the complainant some opportunity to resist him. It is easier to imagine counting his belief that she consented reasonable than counting as reasonable a belief on the part of the California defendant that the sleeping woman (who, in case you are wondering, had **not** said, 'I'm going to sleep, but feel free to penetrate me in my sleep if you want to') was consenting. [In that case, *People v Iniguez*, 7 Cal. 4th 847, 30 Cal. Rptr. 2d 258, 872 P.2d 1183, 1186 (1994), the basis for his appeal was the force requirement: he denied that he had forced her to have sex. The appellate court accepted his argument; the California Supreme Court, I am pleased to say, did not.]

[106] http://www.legislation.gov.uk/ukpga/2003/42/part/1/crossheading/rape/section/1
This Act governs England and Wales, and parts of it govern Northern Ireland; the Sexual Offences (Scotland) Act (2009) has similar language regarding step-taking.

[107] Relatedly, it is worth bearing in mind that the bar for capacity to consent might appropriately be higher for some activities than for others. E.g., there may be cognitive impairments that would render one not capable of giving informed consent to taking part in a medical experiment (at least not without someone with exceptionally good communication skills and no vested interest in how the would-be subject decides

too impaired *cognitively* to be able to consent to sex, he has (or at least in the particular circumstances, had) no way to communicate consent or lack thereof.[108] I will not address the question of whether he in fact *was* capable of consent; my sense is that he was not (because of cognitive impairment), but I am not concerned to defend that position. Rather, I will assume for *the sake of discussion* that the sexual activity was nonconsensual (nonconsensual due to an incapacity to consent), and I'll focus on the question of *whether we should consider Stubblefield's belief that he was consenting to be unreasonable*. I will also assume that Stubblefield really did believe he was consenting. (And I think she did, though there are aspects of the story that give me pause.)[109]

Now, to the details of the case itself, starting with D.J. (I am taking this from the *New York Times Magazine* article by Daniel Engber, "The Strange Case of Anna Stubblefield.")[110] This much is uncontested: D.J., 30 at the time that the alleged rapes took place, cannot speak and never has--in fact has never uttered a word. He screams when he is unhappy and chirps when he is excited. He has trouble making eye contact and keeping objects fixed in view. He can walk only if someone steadies him, and otherwise he gets around by scooting on the floor. {(Why he does not use a wheelchair or a walker isn't clear to me.)[111]} The physical impairments, and the diagnosis of cerebral palsy, are not in question. The psychologist who, in his capacity as consultant for the New Jersey Bureau of Guardianship Services, assessed D.J. in 2004, concluded that

---

engaging in a non-hurried conversation with her about what it involves), yet would not render one not capable of sexual consent.

[108] Having pointed out these two related but distinct bases for claiming someone could not consent, I want to interject that there would be something very wrong if for those whose impairments seriously impede communication (including writing and typing) but do not seriously impede--at least do not rule out--their choosing to engage in sexual relations, the law left it that because of the difficulties in communication, sex for them will have to count as nonconsensual. (This could be a reason for rejecting my position that we should distinguish between not wanting to do X and not consenting to it; but given the dangers of deciding for impaired persons what they want to do, it seems wiser to broaden our notion of what counts as consenting, recognizing that different steps may be appropriate for ascertaining whether someone consents, depending on that person and the particular circumstances.) Relatedly, it is important to recognize that persons whose impairments may make them unable, say, to dress themselves, might be interested in, and in a position to enjoy, sex. The appropriate worries that they will be taken advantage of by others should not be allowed to preclude the possibility of sex for them--more precisely, to preclude the possibility that they could engage in (what is deemed by the criminal law to be) *consensual* sexual activity. (Just how the danger that they will be taken advantage of should be balanced against the need to recognize that they may be in a position to choose, and enjoy, sex is beyond the scope of this paper--and outside my areas of expertise.)

[109] Having read the paper published under his name, written via "facilitated communication" provided by Stubblefield, I find it mind-boggling that she could really have thought that the paper was his work. She surely should have thought to herself that someone who never went to school, was not home-schooled, and could not educate himself, could so quickly move from not talking, reading, or writing to producing a publishable paper. The only possible explanation would be that these words they typed together using facilitated communication were hers, and not his. (Another consideration is that this paper sounds quite a lot like the papers she knows she did author.)

[110] Daniel Engber, "The Strange Case of Anna Stubblefield," *New York Times Magazine*, Oct. 20, 2015.

[111] A disability advocate, Julie Equality, who sat in on the trial, commented that it was troubling that instead of using "a wheelchair, walker, or crutches," D.J. "was physically supported by his mother." "He looked like a baby being guided to take his first steps," observes Equality. Quoted in David M. Perry, "Sexual Ableism," *Los Angeles Review of Books*, Feb. 25, 2016. https://www.lareviewofbooks.org/article/sexual-ableism/

although his impairments precluded any formal testing of intelligence, “his comprehension seemed to be quite limited,” “his attention span was very short,” and--crucially for the purposes of this paper--he lacked “the cognitive capacity to understand and participate in decisions.” The psychologist also reported that D.J. could not carry out basic, pre-school level tasks (presumably for reasons other than only his physical impairments).[112] It is the psychologist's assessment that Stubblefield denies.

{Before I say more about the facts of the case, I want to issue a disclaimer. I am no expert on this case; what I know is taken primarily from two *New York Times Magazine* articles (both by Daniel Engber), together with what I've learned from reading about facilitated communication (statements by advocates as well as a very helpful article by a former facilitator who, after causing immeasurable harm thanks to her trust in it, is at pains to warn against it).[113] (I read a relevant paper of Stubblefield's, and the paper published under D.J.'s name, as well.)}

Stubblefield was convinced that D.J. was a very intelligent man who was merely unable to communicate his thoughts through any of the usual ways (talking, signing, writing, typing on his own).[114] Through facilitated communication she sought to enable him to communicate, and she believed--fully believed--she succeeded. (The method she employed involves holding one hand under the person's elbow, the other over the person's hand, with the facilitated person's index finger extended. The assistance is intended not to guide him or her, but only to facilitate, a facilitation necessary because of problems of motor control and coordination.) They spent a great deal of time together, engaged in facilitated communication and--as she saw it--they fell in love. Sexual connection was challenging because of his physical handicaps, but eventually they (or she) got it to work. She was so convinced that the words that he typed (with her hand over his) were his, including his expression of interest in a sexual relationship with her, that she had no doubt at all that everything they were doing was mutual, including having sex.[115]

From my description so far, and on the assumption that D.J. was not consenting and that the words he typed were hers and not his, Stubblefield seems to be acting out a fantasy (without any notion that it is a fantasy). She imagines him to be deep, thoughtful, full of ideas he is eager to share; she rescues him from a life in which his thoughts were trapped inside him; they fall in love, and now she gives him even more: not just a way to communicate his thoughts but a romantic and sexual relationship. Insofar as we think of her as living out a fantasy, she bears some striking similarities to Benigno, a character in the film, "Hable con Ella" ("Talk to Her"), directed by Pedro Almodóvar.

Benigno, a nurse, has a crush on Alicia, a young dancer whom he observes from afar (thanks to having a view from his apartment of her dance studio). When a car accident

---

[112] This paragraph draws heavily from Daniel Engber, "Strange Case."

[113] Boynton, "Facilitated Communication--what harm it can do: Confessions of a former facilitator," *Evidence-based Communication Assessment and Intervention* 6:1 (2012): 3-13].

[114] Her view, as she explains in the letter she wrote to the judge prior to sentencing, is that she and D.J. are "intellectual equals." (Quoted by Daniel Engber in "What Stubblefield Thought She Was Doing," *New York Times Sunday Magazine*, Feb. 3, 2016. http://www.nytimes.com/2016/02/03/magazine/what-anna-stubblefield-believed-she-was-doing.html)

[115] And judging from her letter from her prison cell to the judge, pleading for mercy, she seems not to have wavered in her conviction that it was consensual. See note 33, above.

seriously injures her, Alicia ends up with Benigno as her nurse. Tending to her in her stable, comatose state goes on for months, even years. Thanks to another nurse needing some time off, Benigno finds himself alone with Alicia for long stretches, including some nights. As viewers, we aren't quite sure what to think about his practice of talking to her as if she can hear and comprehend what he is saying (and not just a few words, but detailed reports of films he has seen); nor do we know what to think of the intimate massages he lovingly gives her, because although they are better for her than being touched more minimally, they are also clearly more gratifying to him than is appropriate. Should we bracket this, we wonder? (A fellow nurse is worried too, though unlike the viewer, she doesn't know of Benigno's longstanding romantic obsession with Alicia.) Is this just a way to make a hard job more enjoyable, harmless and quite possibly good for her (since sometimes a person who appears fully unconscious hears more than we think)? Treating a comatose person as if she can listen to him, telling her in detail about theatre performances, taking her out on a balcony to enjoy the breeze and the sunshine--all this seems better, we tell ourselves, than treating her as just a physical body. But we soon realize that this is not a case of treating her "as if...." As he sees it, they are a couple. When he tells his friend, Marco, that he plans to marry Alicia--not just that he hopes to marry her if she ever emerges from a coma, but that he plans to marry her in her current state--he seems fully unprepared for Marco's reaction of shock and horror. Soon we see how far Benigno has taken what we, but not he, see as a fantasy: she is pregnant.

One of the fascinating features of "Talk to Her" is that Benigno intensely enjoys his imagined relationship in part because it--and to a considerable extent, Alicia--are his own constructions. (Of course, it matters that he doesn't see them as such.) He doesn't have the challenges of a real relationship; he does all the talking and never suffers the hardships that those with real relationships have--feeling put down, being challenged or contradicted when you wanted support, realizing that your partner wasn't listening to you and was bored by your story. He doesn't have to contend with grumpiness, nor worry that something he says will offend her. He can idealize her without having to face disappointment. There are no arguments. "Why shouldn't we get married?" he says to Marco. "We get along better than most married couples."

It seems likely that Stubblefield had something similar (though she presumably had to contend with grumpy moods) but in at least one respect better: she got to create (seemingly real) conversation, giving D.J. his lines. (And of course all the while she fully believed that the words he typed, with her hand on his, were his own thoughts). Benigno talked to a motionless, unresponsive Alicia; by contrast, when Stubblefield declared her feelings for D.J., he typed back "I love you, too." And soon after that: "So now what?"[116] Whereas Benigno had no reply to Marco's emphatic "It's just a monologue!", Stubblefield fully believes that she and D.J. are engaged in dialogue. (And she has a partner who says just what she wants to hear, yet at the same time is an unfolding personality, full of [what seem to her] surprises.)

The jurors reportedly were baffled: how could Stubblefield love him? I think that betrays a lack of imagination, and a failure to appreciate the powers of the imagination and the draw of a fantasized relationship (at least as long as the person immersed in the

[116] Engber, p. *.

fantasized relationship can see it as not a mere fantasy). One can imagine the person to be [almost] whatever one wants him or her to be; one can enjoy being with this loved one without being challenged or contradicted. (And if she wants to be challenged, she can type his lines accordingly.) Yet at the same time, one is not alone, as one is in a purely fantasized relationship. In the case of both Benigno and Stubblefield, there *is* this real person, of real flesh and blood, and the person is (physically) very present. Moreover, this person is one's project. Stubblefield devoted much of her free time to enabling D.J. to communicate, to realize his potential, to live a real life. Thanks to her, he has transformed from someone whose pleasures, apart from eating, consisted (according to Engber's report) mainly in playing with plastic coat hangers into a man who reads voraciously such works as those of Maya Angelou and who, with Stubblefield's help, writes papers that are delivered at conferences. No wonder she loves him!

Enough on the similarities between Benigno and Stubblefield, and between the relationships they create for themselves. Here is the difference I want to highlight: whereas any suggestion that Benigno believed on reasonable grounds that the comatose Alicia consented to sex with him would be utterly preposterous, a suggestion that Stubblefield believed on reasonable grounds that D.J. consented to sex with her is not preposterous.

5. Some Grounds for Deeming Stubblefield's Mistaken[117] Belief Reasonable

Stubblefield was not simply creating her own fantasy in thinking that D.J. was mentally far sharper than the psychologist had determined him to be, and that he was indeed communicating his thoughts through facilitated communication (hereafter, FC). Her faith in FC as a method of communication that generally works (but can go awry if the facilitator is not careful) was shared by many others; her belief that FC was enabling D.J. to communicate his thoughts was held (at least for several months) by D.J.'s mother and brother and by many others. To get into the details, we need to go over the history of Stubblefield and D.J.'s relationship.[118]

Stubblefield and D.J. met in 2009 through his brother, Wesley, a student in one of her classes at Rutgers-Newark. After she showed the class part of a documentary[119] depicting a nonverbal girl with disabilities and an I.Q. of 29 who, thanks to FC, managed to go to college, Wesley told Stubblefield about his brother, and asked if D.J. might be able to utilize FC. Soon she was working with D.J. every other Saturday at Rutgers. Wesley and D.J.'s mother, P., were delighted by his (apparent) progress, and P. invited Stubblefield to her home for more frequent FC sessions for D.J.

Some months later Stubblefield's mother, Sandra McClennan, suggested that D.J. might write a short conference paper for a panel she was organizing for the Society for Disability Studies. Stubblefield and D.J. worked together on the essay, and in June 2010, D.J. traveled with Wesley and their mother to the conference, where Wesley presented

[117] Reminder: I am assuming for the sake of discussion that her belief was mistaken (and I think it very likely was).

[118] Again, I take this entirely from Engber's "The Strange Case of Anna Stubblefield."

[119] "Autism is a World" (2004), written by Sue Rubin, produced and directed by Gerardine Wurzburg, and co-produced by the CNN cable network.

the paper. Subsequently Stubblefield helped D.J. write another conference paper. Stubblefield, D.J., and D.J.'s mother traveled together to this conference, where Stubblefield's father presented the paper. The paper was subsequently published in a peer-reviewed journal, *Disability Studies Quarterly*.[120]

In Fall, 2010, D.J. sat in on a 400-level course in African-American literature, assisted in his homework by FC provided by Sheronda Jones, an undergraduate recruited by Stubblefield. By the time Stubblefield and D.J.'s relationship had taken a sexual turn--Spring 2011--Wesley had begun to have doubts about FC, though he seems not to have relayed these to Stubblefield. Whether P. also had doubts at that time isn't clear, but she and Wesley were frustrated by the fact that unlike Stubblefield, Stubblefield's mother, and Sheronda Jones, they invariably failed when they tried to assist D.J. at typing, despite having spent hours training in FC.

I've recounted all this to offer reasons for thinking that Stubblefield's belief that D.J. was communicating his own thoughts via FC should count as reasonable for purposes of criminal law--not merely understandable, but reasonable. A quick note, though, on reasonableness: I take 'reasonable' to mean only 'not unreasonable' and take this to be the case both for the term as we use it in ordinary conversation and as we use it--or should use it--for criminal law purposes.[121] It should mark a threshold, and not a very high one given what is at stake (a prison term). {In another context one might use the notion to mark a high achievement, but this cannot be correct in the context of criminal law.}

I take it to be evidence (but certainly not conclusive evidence!) that her belief was reasonable that those who knew D.J. best also saw him to be conveying his own thoughts and that she knew this. {(As noted, their confidence--at least Wesley's--was slipping by 2011, but it appears unlikely that this was conveyed to Stubblefield.)} The student who assisted him in the course he sat in on seems to have had no doubts. Stubblefield's work with D.J. was also affirmed by her parents. From what I have read, no one was relaying to Stubblefield any worries about whose thoughts were being typed out, hers or D.J's, and her confidence in him and in FC seems to have been reinforced by a good-sized circle of people: her parents,[122] D.J.'s mother and brother, those who attended the conferences where his papers were presented, and (arguably) the editor(s) of the journal where one of the papers was published.[123] It was bolstered yet more by the community of FC providers. Her belief in FC and her facilitation skills were lent further support by the appreciation of others for whom she served as a facilitator, and their parents.[124]

---

[120] http://dsq-sds.org/article/view/1717/1765

[121] I elaborate on this in "Reasonableness," unpublished manuscript.

[122] Probably also her brother. He testified at the trial that had she and D.J. married, as she had intended, the family would have welcomed him into the family with open arms.

[123] From the content of the paper it is made pretty clear that the author is writing the paper using facilitated communication; moreover, the contact information provided for him is Stubblefield's. That they accepted the paper thus seems to be an endorsement of it as an outcome of facilitated communication--though to be sure, they would be unlikely to know the exact nature of the facilitated communication, since this varies depending on the capabilities and needs of the FC user, and thus might not realize how much of a risk there was that the facilitator was steering him or her.

[124] Engber mentions that one of them, Zach DeMeo, came with his mother to Stubblefield's trial to show their support. Zach's mother is quoted as saying, "It changed his life. She was so selfless and devoted...She

If her belief that he was communicating his thoughts through FC was reasonable, so was her belief that the psychologist's assessment was totally wrong, and that D.J. was not mentally handicapped. After all, he could write conference papers, and fruitfully sit in on an advanced undergraduate course! And if he could do all that, what basis would there be for saying that he was incapable of consenting to sex? On the assumption that FC was working for D.J., the reasons for thinking either that he does not have the "mental age" to consent or that he cannot communicate consent or lack thereof evaporate.

To be sure, many of the people who believed FC worked--specifically, that it worked for D.J.--believed this not on independent grounds, but because Stubblefield believed it. Wesley learned about FC from her, as did his mother; Sheronda Jones, the student who assisted D.J. in the course he audited, was no doubt influenced by the views of Stubblefield, who recruited her to aid D.J. (Whether she antecedently believed in the reliability of FC, I don't know.) But there was much more ratification than just that. Her parents did not support her merely out of friendly support for a daughter; Stubblefield's mother, a Ph.D. in Special Education who had been working with cognitively impaired children since 1963, began using FC long before Stubblefield did. She is regarded as an expert on FC and continues to advocate for it. I mentioned confirmation from the community of FC providers, and we should bear in mind that that community is not on a par with, say, palm-readers or astrologers. The main institute for FC is housed at Syracuse University (though in 2010 it changed its name from 'Facilitated Communication Institute' to 'Institute on Communication and Inclusion' because of controversy about FC).[125] Although highly controversial, FC has many academic supporters. In addition, FC is endorsed by the Autism National Committee in a policy statement.[126] {Also worth mentioning as providing some support for Stubblefield's confidence in FC is that the film mentioned above (part of which she showed to the class of hers that Wesley took), "Autism is a World," was nominated for an Oscar.}

I mentioned earlier that the steps taken to ascertain whether the other party is consenting should factor into an assessment of the reasonableness of the belief. So we should note that Stubblefield (at least according to her testimony)[127] sought to ensure ongoing communication from D.J. during their sexual encounters. If D.J. needed to say something, he would bang on the floor, and she would pause to set him up with the keyboard.[128] {She thus contrasts to Tawera, who did less to ascertain whether his "partner" was consenting.[129]}

---

speaks to my son as an equal. She treats him as a human being. If he told me he was in love with her, I would believe him" ("Strange Case," p. 19).

[125] Engber, "The Strange Case of Anna Stubblefield," p. 22.

[126] http://www.autcom.org/articles/PPFC.pdf

[127] I am assuming throughout my discussion that what Engber reports her as having said she said sincerely; I do so because I am interested in the question of whether her belief, given the facts as presented, was reasonable, not whether she reported anything that she did not take to be true. That said, it is worth bearing in mind that all the evidence against her came from her. If she were going to try to help herself out by being less than fully truthful, she could easily have done a far better job! For starters, she could have refrained from informing D.J.'s mother and brother that she and D.J. now had a sexual relationship.

[128] Engber, "Strange Case," p. 18.

[129] The contrast is even greater to more typical rape cases, where the defendant showed no concern at all in whether the other party was consenting. See for example the case described above in note 24.

6. Some Grounds for Thinking Stubblefield's Mistaken[130] Belief Unreasonable

But there is also ample room for doubt about the reasonableness of her belief that D.J. was consenting. She held (and, from the last report I have, holds)[131] her beliefs about the reliability of FC and its effectiveness with D.J. with a fierce tenacity, refusing to consider the possibility that what D.J. typed were her thoughts, not his. This to my mind is the main (and probably the only) reason for thinking that her belief that he was consenting was not reasonable or, to put it in the terms used in the N.J. statute under which she was charged and convicted, that she "should have known" that he wasn't consenting.[132]

Even if we assume that (prior to her announcement that they were in love) no one directly challenged her assumption that FC was working properly for D.J., there is no question but that Stubblefield was aware of the possibility of a facilitator unknowingly guiding the typing. The case of Betsy Wheaton, an autistic teenager whose parents were charged with child abuse solely on the basis of what she and her facilitator typed using FC, had been widely publicized in 1993 on "Frontline" and in 1994 on "20/20." Many other such cases came to light, and as FC was tested (by, e.g., asking the FC user questions the answers to which the FC facilitator wouldn't know, but which the FC user should know), it became clear that the "ideomotor" effect was extremely common, and that FC (in the form that Stubblefield used)[133] very rarely (if ever) worked.[134] The American Psychological Association issued a resolution in 1994 that there was "no

---

[130] Reminder: I am assuming for the sake of discussion that her belief was mistaken (and in fact do think it very likely was).

[131] In her letter to the judge, pleading for mercy, she makes it clear that she has in no way revised her view that D.J. was "the author of his words and a very intelligent man." Quoted by Engber in "What Anna Stubblefield Believed She Was Doing."

[132] Or more aptly, though less like the language of the NJ statute: should have suspected that he might not be consenting.

[133] I say this because 'FC' may also be used to refer to facilitation that involves far less guidance, e.g. steadying an elbow or holding the keyboard. In addition, it can be used briefly, to enable the user soon to type independently. The (particularly/clearly) problematic cases are those where there is no transitioning to independent typing and the facilitator has his or her arm and hand over the user's. (Note: some would say that all uses of it are problematic. I do not mean to be denying that, but do not know enough to affirm it, though the instances where the users transition to independent typing provide reason to think it may have some value. See Perry, "Sexual Ableism," cited above in note 30).

[134] Janyce Boynton, a former facilitator, says "Every facilitator moves their communication partner's arm and authors the FC messages" (Boynton, "Facilitated Communication," p. 12). However, her position is not entirely clear, because two sentences back she says "[I]f I were a school administrator, educator, parent, caregiver, guidance counselor, lawyer, DHS worker, police officer, or judge, knowing what I know today about FC, I would not allow a single word to be typed on a keyboard on behalf of a child without first testing the facilitator in a controlled environment away from the supportive gaze of other believers." That sounds right to me, but it doesn't make a lot of sense if she thinks that invariably the facilitator moves the communication partner's arm and authors the FC messages, for if that is the case, why allow it at all? (Or does she mean only that every facilitator at least sometimes does that?)

scientifically demonstrated support for its efficacy."[135] For those serving as FC facilitators, there was no escaping the claims that FC was at best unreliable, at worst completely worthless. No escaping--but that doesn't mean they gave them serious consideration.

The evidence that Stubblefield knew of the controversy about FC and knew of the false accusations of sexual abuse {(more accurately, the false belief, leading to accusations, that the FC user was reporting sexual abuse)} comes not only from it being impossible for her not to know, but from her published work. In her "Sound and Fury: When Opposition to Facilitated Communication Functions as Hate Speech," Stubblefield dismisses the worries that FC is unreliable, pointing out that this is true of other modes of communication, too, so why all the focus on FC? [136] Her suggestion is that there is great resistance to recognizing that those who appear to many of us to be mentally disabled are in fact often only physically disabled, and that when they are shown respect, encouragement and most of all, faith in their abilities and an interest in hearing what they have to say, they show themselves to be far more intelligent than their IQ assessment suggests. "[T]o an observer who assumes that the FC user is profoundly intellectually impaired, it will appear unbelievable that he can be given access to a means of communication that involves literacy and immediately type meaningful words and sentences."[137] "Anti-FC expression functions as hate speech when it calls into question, without substantiation, the intellectual competence of FC users, thereby undermining their opportunity to exercise their right to freedom of expression."[138] As for the scientific research, in addition to questioning what it really establishes, Stubblefield endorses the following statement, by another author: "Research is really useless as its own reward. The only good purpose for research is liberation from our limitations. Research designed to make those limitations more real and more legitimate must be stopped."[139]

It is clear from her published work that she was well aware of the FC controversy. It is also evident that she had no interest in considering the possibility that FC might be unreliable, and a great deal of interest in discrediting the objections to it. This supports the position that her belief in FC was unreasonable, and likewise her belief that D.J. was consenting to sex with her. But the matter is very complicated. Here are several considerations that should be taken into account as we think about whether her belief that D.J. consented to sex with her was unreasonable.

1. The bar for reasonableness for purposes of the criminal law needs to be fairly low. When we speak of a reasonable belief in the context of a discussion in epistemology, the bar is higher. In such a context we would have no hesitation about saying that her belief that FC was reliable was not a reasonable belief. But should we say the same when we are talking about a reasonable belief for purposes of assessing whether the person meets the mens rea requirement for a conviction? It is not obvious to me that we should. It

[135] http://www.apa.org/research/action/facilitated.aspx. A similar stand was taken by the American Academy of Pediatrics. http://www.therapiesonore.net/wp-content/uploads/2013/07/AuditoryIntegrationTraining1.pdf

[136] *Disability Studies Quarterly* 31:4 (2011). http://dsq-sds.org/article/view/1729/1777

[137] *Ibid*.

[138] *Ibid*.

[139] Eugene Marcus, reportedly an FC user. Quoted in Stubblefield, "Sound and Fury."

seems to me that for such purposes the most important consideration is what steps the defendant took to be sure that the person was consenting.[140]

2. In assessing reasonableness (whether for purposes of criminal law or when the bar is higher), it matters whether the controversial background beliefs upon which the belief in question rests are fairly widely held. That they are controversial certainly need not rule out the possibility that the belief that is based on them was reasonable. That said, if they are held by a large number of people, that may not be enough. Other considerations enter in, to wit:

3. Are the controversial supporting beliefs held only by a very insular, us-against-them community?

4. Is there a way to test a (crucial) supporting belief, and if there is, at what cost? If the defendant did not opt to have it tested, despite it being low cost or cost-free to do so, why not?

5. When, as in this case, the background beliefs are held tenaciously, what is the underlying motivation? Arguably it matters whether the motivation for the belief is (e.g.) to improve the lot of others or (e.g.) to provide oneself with a rationalization for exploiting or abusing others. (One might, however, say that this should factor in only at sentencing, not for purposes of assessing reasonableness, and thus for determining whether the elements of the crime have been proven beyond a reasonable doubt.)[141]

I said earlier that I think the main, and probably the sole, basis for deeming Stubblefield's belief that D.J. was consenting unreasonable[142] is that she refused to consider the evidence against the background belief on which her belief that he was consenting rested. Relatedly, although she did make an effort to ensure that he was consenting by providing him with the keyboard whenever he had something to say, as well as engaging him in (what she saw to be) conversation about their relationship,[143] she failed to do something which she absolutely should have done. She failed to have her use of FC with him tested. The test is simple, painless, and cost-free {(unless of course one finds out that FC is not working, but the possibility of that cost obviously should not be factored in deciding whether to be tested)}.

Is this failure enough to render her belief that he was consenting unreasonable? Quite possibly. But I find it hard to say, for two reasons. First, a detailed account from a former FC facilitator, Janyce Boynton, brings out how difficult it would be for someone in the FC community, particularly someone as invested in it as Stubblefield was, to opt to

---

[140] With regard to other crimes, it won't be consent, but a different factor to which the defendant should be attending. The Model Penal Code definition of the culpability level of negligence is useful to bear in mind: "A person acts negligently with respect to a material element of an offense when he should be aware of a substantial and unjustifiable risk that the material element exists or will result from his conduct. The risk must be of such a nature and degree that the actor's failure to perceive it, considering the nature and purpose of his conduct and the circumstances known to him, involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation" (Model Penal Code 2.02).

[141] The relevant element is the mens rea component: that she knew or should have known that he wasn't consenting; or, framed as I've been framing it, that it isn't the case that she believed on reasonable grounds that he was consenting.

[142] Reminder: I am assuming throughout that he in fact was not consenting, and asking whether, granted that assumption, her mistaken belief that he was is a reasonable belief.

[143] Throughout my discussion I am, as noted above, assuming the veracity of her testimony.

be tested.[144] The pressures against doing so were enormous; FC was supposed to require trust in both the person one was helping and the process itself; also worth noting, and emphasized by Boynton in her confessional report, is the fear of learning that one is one of the bad facilitators who do it improperly and give FC a bad name. None of this justifies Stubblefield's failure to have her use of FC with D.J. evaluated, but it does suggest that it would be setting the bar for reasonableness (for purposes of criminal law) a bit on the high side to deem it unreasonable, given that it would be a very rare and unusually courageous FC facilitator who would opt for it.

Second, the fifth item above does give me pause. The ideology that motivated Stubblefield's (unwarranted) confidence that the words typed were D.J.'s, not hers, was the conviction that one should err in the direction of overestimating, not underestimating, the capacities of the person thought to be cognitively disabled. We do best (the ideology has it) to assume that the person's mental age matches his chronological age unless the evidence forces us to revise this; we do best to figure that the impairments are only physical; we do best to figure he has thoughts, wants to learn, wants to live as an independent adult, and try then to facilitate that. Are we to wait, one might ask, until either a more effective method is found, or FC is determined to be pretty reliable after all, when we have people leading extremely limited lives who might be helped by FC to express their thoughts, become more independent, gain more control over their lives? This ideology certainly has the potential to cause great harm; it is obviously not innocuous. Those who adopt it need to employ safeguards lest it do so. But my point is that the tenacity of her beliefs and her unwillingness to take criticisms of FC seriously were not due to an ugly motivation such as that of Clifford's shipowner.[145] They seem to be motivated by a genuine concern to enable those with disabilities to lead richer lives.

Stubblefield had a long history of social activism, reflected both in her research and in her volunteer work. There is no evidence that anything she was doing with D.J. or with the others she sought to help was self-serving (apart from being self-serving in the way many a project serves to boost the agent's ego), far less that it was predatory. It was based on a fantasy and an ideology, but a fantasy shared by others (referring now to the fantasy that he was much more intelligent than he had been assessed to be) and an ideology that was not only firmly held by a large number of people (and such organizations as the Autism National Committee) but also reflected laudatory goals and arguably admirable attitudes towards people with severe disabilities.

One might say that this does nothing to undermine the claim that she should have known, that she is culpable for not having known, or that she acted unreasonably in taking it that he was deeply intelligent, not cognitively impaired at all, and capable of sexual consent. I am not sure, and look forward to your thoughts on this. But it bears emphasis that I am considering these questions in the context of criminal law, not in the context of a moral appraisal of her conduct. If I were simply engaging in a moral appraisal, I would have no hesitation to say that she should have known (something I

[144] See note 33, above. It was Boynton's "facilitating" that led to the arrest of Betsy Wheaton's parents or child abuse.

[145] I'm referring to W.K. Clifford, "The Ethics of Belief" in *The Ethics of Belief and Other Essays* (Amherst, New York: Prometheus Books, 1999).

have no hesitation to say in any case), that she is culpable for not having known, and that she acted unreasonably in taking it that he was capable of sexual consent.

7. A Possible Objection

I indicated that I thought that Stubblefield's refusal to consider evidence against her belief that he was consenting (in the form of evidence against the assumptions that serve as a foundation for her belief that he was consenting) was probably the only basis for considering her belief unreasonable. I had in mind in particular her failure to take necessary steps to ascertain whether he was consenting, in particular, the crucial step of having her use of FC with him tested.

One might argue, however, that given what I said about Tawera, I should take the position that Stubblefield's belief that D.J. was consenting was clearly unreasonable. I said that Tawera's belief that his cousin was consenting to sex was based only on her passivity, and that because it is explicitly stated in NZ law that passivity alone does not constitute consent, his belief should not count as reasonable. One might claim that for similar reasons, Stubblefield's belief that D.J. was consenting cannot be reasonable because as a matter of law, he could not consent. He was deemed by the state of New Jersey to be mentally below--far below--the age of an adult, and therefore was appointed guardians, and Stubblefield knew this. Hence she either knew or should have known that he was as a matter of law incapable of consent.[146] If she judged otherwise, that was a mistake of law, just as (I claimed) Tawera's view that his cousin was consenting involved a mistake of law.

I don't think this is correct. That consent is defined as *X* is a matter of law; that person *S* is incapable of consent (consent as defined by the law) is not a matter of law. The psychologist could have gotten it wrong. Evidently a great many people thought he had, including (until late on) D.J.'s guardians. If you think *S* has the mental age of a toddler, you do not endorse, as his guardians did, his writing conference papers and sitting in on (or taking) college courses. So I do not think that the fact that the State of New Jersey said D.J. had the mental age of a toddler entails (or even goes a significant distance towards showing) that Stubblefield's belief that he was consenting was unreasonable.

8. A Very Brief Conclusion

There is no question but that Stubblefield acted wrongly. Even if it had been the case that D.J. clearly could consent to sex and was consenting, a sexual relationship was totally off limits because of her role as his facilitator. It would have been off limits for roughly the same reason that a sexual relationship between a dissertation director and her student is off limits. It would not amount to sexual assault.

She also acted wrongly in not considering the possibility that D.J. might be incapable of consenting to sex. Her policy of erring in the direction of overestimating, rather than in underestimating, a person's capacities {(a reflection of the "criterion of the least

[146] This seems to have been the position of the judge. Engber reports that Judge Teare held that Stubblefield "knowingly and wantonly overstepped the bounds of lawful behavior." She "violated the terms of D.J.'s guardianship because she decided, on her own, that the courts were wrong — and that she knew better than the State of New Jersey." http://www.nytimes.com/2016/02/03/magazine/what-anna-stubblefield-believed-she-was-doing.html

dangerous assumption")[147]} needed to be carefully bracketed. Whether her culpability should suffice for the purposes of criminal law--whether we should hold that her belief that he was consenting was unreasonable--is not clear to me. At issue is both how high the standard should be, and also whether the fact that she holds a Ph.D. in philosophy should factor in...to count against her. How, we ask ourselves, could someone with good enough critical thinking skills to get a Ph.D. in philosophy[148] not do a better job at thinking critically? If the bar for reasonableness has to be set low enough that it doesn't require heroism to reach it, should it be raised a notch or two if one has the skills and practice in reasoning that should enable one to rise above the rhetoric about FC and think critically about possible dangers in relying on it?[149]

[147] As put forward in an influential paper, Anne M. Donnellan, "The Criterion of the Least Dangerous Assumption," *Behavioral Disorders* 9:2 (1984), pp. 141-50. The basic idea of erring in the direction of overestimating a person's capacities is picked up on by many practitioners. See e.g. http://www.thinkingautismguide.com/2010/07/living-least-dangerous-assumption.html, where the author asks rhetorically how we go about living the least dangerous assumption, and includes among the answers: "Give the gift of assuming intentionality in communication," explaining that "even if you are wrong in your assumption you will teach intentionality by responding as if the action was intentional." See also Carol Jorgensen, "The Least Dangerous Assumption: A Challenge to Create a New Paradigm," *Disability Solutions*, 6: 3 (2005) and https://inclusivelife.files.wordpress.com/2007/09/least-dangerous-assumption.pdf.

[148] We might add: at Rutgers (main campus), a top-notch philosophy department.

[149] An earlier draft of this paper was presented to the philosophy department of Loyola University of Chicago (2016) and at a conference on sexual consent and coercion at the University of Virginia (2016). I am grateful to discussants at both events for their comments, and especially to Elizabeth Barnes, my commentator at the University of Virginia.

Do I Have to Be Coherent to Be Reasonable?

Alex Schaefer and Wes Siscoe

Abstract: In *Political Liberalism*, Rawls famously denies that his political constructivism needs to reference the concept of truth, a claim that has been criticized by Joseph Raz, Joshua Cohen, and David Estlund. In this paper, we argue that these criticisms fail due to the fact that parties to the overlapping consensus do not have to be coherent in order to be reasonable. Once it is seen that Rawls's political constructivism allows this freedom to reasonable parties, the demands made by Raz, Cohen, and Estlund can be seen to require more of reasonable people than is necessary for a political consensus.

I. Introduction

Central to John Rawls's *Political Liberalism* is an account of political justification, and central to this account of political justification is the method of political constructivism. Notoriously, Rawls asserts that this method of justification functions without recourse to the concept of truth:

"[Political constructivism] does not...use (or deny) the concept of truth; nor does it question that concept, nor could it say that the concept of truth and its idea of the reasonable are the same. Rather, within itself the political conception does without the concept of truth."[150]

Rawls thus holds that, in some sense, the process of justifying political doctrines can do without the concept of truth. This position has been met with mixed reviews. Several commentators think that Rawls's position on the exclusion of truth from political discourse is incoherent:

Joseph Raz —

"To recommend [a theory of justice] as a theory of justice for our societies is to recommend it as a just theory of justice, that is, as a true, or reasonable, or valid theory of justice."[151]

Joshua Cohen—

"The idea of locating a common ground of political reflection and argument that does without the *concept* of truth...is hard to grasp. Truth is so closely connected with intuitive notions of thinking, asserting, believing, judging, and reasoning that it is difficult to understand what leaving it behind amounts to."[152]

David Estlund—

[150] See Rawls, Political Liberalism, p. 94.
[151] See Raz, Facing Diversity, p. 15.
[152] See Cohen, Truth and Public Reason, p. 15.

"Political liberalism must assert the truth and not merely the reasonableness - or acceptability to all reasonable people - of its foundational principle."[153]

The animating concern of all these critiques is that Rawls smuggles in or depends upon the concept of truth all the while claiming to avoid it.

In this paper, we will defend Rawls, arguing that there is a coherent way to understand his comments on the relationship between political justification and truth. The crux of our argument is that reasonable parties to the overlapping consensus can be conceptually incoherent in that they can have contradictory beliefs concerning their most basic concepts. In section 2, we will detail the critiques of Raz, Cohen, and Estlund, arguing in section 3 that all these objections are undermined by the fact that reasonable people can be incoherent. Section 4 contains a fuller response to Rawls's critics before considering some objections in section 5. Ultimately, we conclude that Rawls's political constructivism is vindicated because participants in the overlapping consensus do not have to be coherent in order to be reasonable.

II. Rawls's Critics

1. 3 Criticisms of Truth Avoidance

There have been several criticisms leveled against Rawls's exclusion of the concept of truth from political constructivism. The first of these critiques comes from Joseph Raz, who argues that the acceptability of Rawls's theory of justice is inconsistent with remaining agnostic on its truth. In Raz's view, regarding a principle of justice as acceptable entails regarding the principle as true:

> To recommend [a theory of justice] as a theory of justice for our societies is to recommend it as a just theory of justice, that is, as a true, or reasonable, or valid theory of justice. If it is argued that what makes it the theory of justice for us is that it is built on an overlapping consensus and therefore secures stability and unity, then consensus-based stability and unity are the values that a theory of justice, for our society, is assumed to depend on. Their achievement – that is, the fact that endorsing the theory leads to their achievement – makes the theory true, sound, valid, and so forth. This at least is what such a theory is committed to. There can be no justice without truth.[154]

According to Raz ,if a political conception is acceptable as the focus of an overlapping consensus and if this acceptability vindicates its principles, then the political conception must be "true, sound, valid, and so forth" in virtue of its ability to serve as the focus of an overlapping consensus. In other words, recommending a theory of justice (according to

[153] See Estlund, Insularity, p. 253.
[154] See Raz (1990), p. 15. Emphasis in the original.

any given standard) commits one to asserting that the theory is true. If this is correct and the theory of justice put forth in *Political Liberalism* actually does live up to Rawls's standards, then the theory of justice is also put forward as true. Rawls implicitly claims that his theory is true, thereby failing to avoid the concept of truth as he had hoped to.

In a similar vein, Joshua Cohen sees a contradiction in being non-committal with respect to the concept of truth while still employing other concepts closely related to truth. Many of the activities associated with political deliberation appear to be conceptually connected with truth – activities "thinking, asserting, believing, judging, and reasoning"[155] – and thus it is problematic to employ these concepts in political deliberation while simultaneously eschewing all reference to truth. One of these activities that Cohen explores in more detail is the concept of believing. It is commonly held that beliefs aim at being true, and that insofar as a person accepts the falsity of a proposition, they cease to believe it.[156] If this is correct—and it seems to quite plausible—then by believing that *p*, a person undertakes a mental commitment to the truth of *p*. It is therefore unclear how parties to the overlapping consensus can believe in the accepted conception of justice without simultaneously affirming it as true. For such reasons, Cohen maintains, Rawls's vision of political deliberation cannot proceed without some concept of truth.

Whereas both Raz and Cohen emphasize the apparent incoherence of engaging in public deliberation while rejecting the concept of truth, David Estlund takes a different tack, arguing that the inconsistency in Rawls's thought lies in the avoidance of truth *along with* the claim that Rawls's principles of justice can create actual moral obligations. On Rawls's view, reasonableness can play the role of truth in political deliberation by adjudicating between competing conceptions of justice to regulate the basic structure. Accordingly, Rawls employs reasonableness as a criteria for any doctrine to be included for consideration. The political conception that constitutes the focus of an overlapping consensus attains vindication via its reasonable acceptability, not truth. Against this suggestion, Estlund argues that such acceptance could not ground moral obligations:

> Suppose, in order to avoid the truth, we understand political liberalism not as ordering an account of the true standard but simply as using a standard that is acceptable to all reasonable people (the standard itself being acceptability to reasonable people)...The question is whether it could ground obligation and justify coercion even if the acceptance criterion it uses were not true. Never mind for the moment whether political liberalism says anything on this question; the answer to the question is that it could not have those moral consequences irrespective of the truth on those matters.[157]

Estlund holds that the only way that Rawls's process of political deliberation could yield moral obligations is if the principles governing that deliberation were in fact true. If

[155] See Cohen, p. 15.
[156] See Williams (2002), p. 67
[157] See Estlund (1998), pp. 261-262

Estlund is right, then Rawls cannot maintain his ambivalent attitude towards truth and that his theory of justice leads to moral obligations for the reasonable participants in the overlapping consensus.

To drive his point home, Estlund asks us to consider the following thought. Rawls wants us to believe that what matters in political deliberation is acceptability to reasonable people. But why should acceptance to such a group matter? There are plenty of groups we could choose from— acceptability to all redheads, or to all members of the Star Davidian cult, for instance—but what is lacking is a criteria for selecting one of these groups from among the others. One response, that reasonable people tend to settle on true principles of justice, is not available to Rawls due to his agnositicism about truth. Estlund thinks that without claiming that reasonable acceptability is the "true" standard of admissibility in public discourse, Rawls's "view loses any way to select among the plurality of insular groups, and it becomes untenable."[158] Without specifying why acceptability to reasonable persons is a better standard than acceptability to Star Davidians, Rawls's account fails to justify his particular principles of justice. Insofar as Rawls fails to justify his principles of justice, Rawls also fails to explain why the overlapping consensus gives rise to moral obligations. An agreement amongst all Star Davidians would not give rise to such obligations, so why think that a consensus of reasonable people would?

2. An Alternative to No Concept

All three criticisms of Rawls object to the same, stringent doctrine concerning the role of truth in political justification:

No Concept: Political Constructivism does without appealing to the concept of truth as well as to any concept that is conceptually linked to truth.

Neither Raz, Cohen, nor Estlund think that Rawls can get by with No Concept. Raz thinks that the act of recommending a principle of justice is conceptually linked to truth. Cohen argues that there are several concepts at play in deliberation, like belief, assertion, and reasoning, that are conceptually connected to truth. Estlund holds that moral obligations can only be created by a political foundation that is in fact true. These conceptual connections to truth end up implicating Rawls's Political Liberalism as dependent on the concept of truth.

An attractive alternative to Rawls's truth-abstinence, suggested by both Cohen and Estlund, can be formulated as follows:

[158] Ibid, p. 260.

Concept Indifference: Even though political constructivism appeals to the truth of some claims and concepts that are conceptually connected to truth, it need not adjudicate between competing understandings of truth

The benefit of Concept Indifference is that it can address all of the previous worries without being too exclusive. Rawls could simply agree with his critics that truth does have an important role to play in establishing the structure and role of political discourse yet nevertheless remain agnostic about which is the correct theory of truth, whether that be some version of correspondence, minimalism, pragmatism, etc. This thin concept of truth could then be used to respond to the critiques of Raz, Cohen, and Estlund, or so the story goes.

III. The Cost of Concept Indifference

Between the criticisms of Raz, Cohen, and Estlund, there seems to be good reason to question the eschewal of truth that Rawls advocates. However, to reject Rawls's approach before considering the nature of and reasons for Rawls's abstention from this concept would, of course, be premature. In fact, examining the reasons that Rawls states for avoiding truth reveals an insoluble tension between the recommended approach of Rawls's critics, Concept Indifference, and the conjunction of two key concepts in Rawls's framework, namely the reasonable and full publicity. This section begins by presenting Rawls's motivation for avoiding the concept of truth. We then argue that involving the concept of truth in the procedure of political constructivism would undermine key goals of Rawls's project.

1. *Why Avoid the Truth?*

The problem with Concept Indifference is that it is inconsistent with Rawls's central motivation for avoiding the concept of truth. Throughout *Political Liberalism*, Rawls describes the central problem that he intends to address in various ways, for example: "How is it possible that deeply opposed though reasonable comprehensive doctrines may live together and all affirm the political conception of a constitutional regime?"[159] His solution comes in the form of "political constructivism," a procedure wherein Rawls leverages certain conceptions of the person and of society that he takes to be implicit in the culture of any liberal democratic society. Due to their latent presence in the public culture, such conceptions are--at least implicitly--endorsed by all citizens. In this way, any principles of justice that emerge from these commonly held values could be endorsed by all citizens in the liberal democratic society.

To understand what is meant by "could" one must consider Rawls's notion of reasonableness. Rawls takes the reasonableness of persons to be one of those implicit

[159] *Political Liberalism*, xviii

conceptions in a liberal democratic society. Unravelling the exact concept of "reasonable," however, is no simple task.[160] The two most important notions of the reasonable are (1) reasonable doctrines[161] and (2) reasonable persons. For the topic of political constructivism, it is the second of these two notions that plays a dominant role.

Rawls identifies two key features of reasonable persons. In Lecture III of *Political Liberalism* he writes:

> The idea of the reasonable is given in part, again for our purposes, by the two aspects of persons' being reasonable: their willingness to propose and abide by fair terms of social cooperation among equals and their recognition of and willingness to accept the consequences of the burdens of judgment.[162]

In other words, a reasonable person is a conditional, rule-following cooperator, and tolerates diverse viewpoints in light of the difficulty of coming to conclusions on matters of faith, morality, and other fundamentals. For a reasonable person then, the most attractive political conception of justice will be one that all could recognize as an adequate compromise. Such a conception, Rawls submits, is one that is built up from values that all endorse, values that are not particular to any comprehensive doctrine. Therefore, citizens, insofar as they are reasonable, recognize that a political conception of justice based on commonly held values and conceptions is one that is worthy of endorsement. Such a conception could thus operate as the focus of an overlapping consensus among reasonable persons.[163]

Although this brief sketch leaves out most of the richness (and all of the deep difficulties) of Rawls's approach, it helps to clarify why Rawls sought to avoid appealing to the concept of truth. If citizens endorse diverse comprehensive doctrines, then insofar as the task of a political conception of justice is to reconcile these disparate doctrines, to gain unanimous assent, to function as the focus of an "overlapping consensus" -- to this extent, a political conception of justice must aim to be neutral with respect to divisive philosophical or religious commitments. One such commitment is the nature of truth and its relation to normative, e.g. political, propositions. Since, by hypothesis, all citizens are committed to achieving a mutually acceptable conception of justice, using the reasonable

---

[160] See Leif Wenar (1995).

[161] "They [reasonable doctrines] have three main features. One is that a reasonable doctrine is an exercise of theoretical reason: it covers the major religious, philosophical, and moral aspects of human life in a more or less consistent and coherent manner… In singling out which values to count as especially significant and how to balance them when they conflict, a reasonable doctrine is also an exercise of practical reason… A third feature is that while a reasonable comprehensive view is not necessarily fixed and unchanging, it normally belongs to, or draws upon, a tradition of thought and doctrine" (PL 59).

[162] *Political Liberalism,* 94.

[163] I ignore the meaning of reasonable in the context of "reasonable comprehensive doctrines," because this notion is far too permissive to achieve what Rawls wants and, furthermore, was dropped by Rawls in his later essay: "Public Reason Revisited"

as a standard by which to test conceptions of justice is less divisive than using truth as a standard. As Rawls writes, "One thought is that the idea of the reasonable makes an overlapping consensus of reasonable doctrines possible in ways the concept of truth may not."[164] Various comprehensive doctrines may not countenance the concept of truth, and avoiding the concept of truth in political constructivism allows those who hold such doctrines to endorse the focus of the overlapping consensus without contradicting their own particular comprehensive doctrines.

Closely connected to political constructivism, reasonableness, and the idea of unanimous acceptability (the overlapping consensus) is a desideratum that Rawls calls the "publicity condition."[165] A society can satisfy the publicity condition on three distinct levels of increasing demandingness:

> 1) Citizens know and accept a single conception of justice. In addition, they accurately and justifiably believe, as a part of common public knowledge, that society's institutions satisfy the demands of this conception of justice.
>
> 2) Citizens affirm the same empirical, social facts that are relevant to political justice.
>
> 3) The full justification (i.e. the argument in support of) the political conception of justice is publicly known or publicly available.[166]

Rawls cites several reasons for the importance and desirability of satisfying the publicity condition. However, to appreciate this condition on an intuitive level, it suffices to recall the aim of political liberalism and the function of political constructivism. As mentioned above, *Political Liberalism* seeks to offer an account of how "a plurality of reasonable doctrines, both religious and nonreligious, liberal and nonliberal, may endorse" a single political conception of justice. To this end, Rawls employs a constructivist procedure that draws solely from society's stock of shared values and conceptions, namely those that are implicit in the public political culture. All three levels of publicity concern the understanding and endorsement of the political conception, its realization, and the reasons that underlie and justify it. Without satisfying all three levels, some citizens in such a society cannot fully, cognizantly endorse the governing political conception of justice. Therefore, satisfying the publicity condition is necessary for fully realizing the aim of *Political Liberalism*. And furthermore, the method of political constructivism is devised as a means of making the political conception understandable and justifiable to the citizenry as a whole -- i.e. as a means of satisfying the publicity condition. It is this desideratum that drives Rawls to avoid truth, and to instead employ the public conception of reasonableness as the standard by which to judge a political conception of justice.

---

[164] See Rawls (2005), p. 94.

[165] *Ibid.* 66. We would like to thank Brian Kogelmann for calling our attention to the relevance of this aspect of Rawls's project.

[166] *Ibid.* 66-7

Truth, being denied or doubted by many reasonable citizens, appears to be inconsistent with the ideal of publicity, and therefore with the aim of *Political Liberalism*.

2. Two Concrete Cases

We have struck upon a certain inconsistency between two approaches to political justification: (a) involving the concept of truth in the procedure of political constructivism, and (b) satisfying the publicity condition for all reasonable doctrines in a liberal democratic society. To concretize this inconsistency, we present two cases of reasonable persons endorsing reasonable doctrines that cannot consistently accept a political justification that employs the concept of truth.

Consider the following doctrine:

Political Noncognitivism: Political propositions are neither true nor false.

There are perfectly sane, cooperative people who endorse such a doctrine. Certainly, persons endorsing such a doctrine *could* be reasonable: they could be conditional cooperators who recognize the burdens of judgment. This doctrine may be part of a comprehensive doctrine that Rawls would call "reasonable." Hence, if Rawls required that all the comprehensive doctrines included in the overlapping consensus regard political propositions as true, then a resonable view held by reasonble persons (including some of our colleagues) would be excluded.

Consider a second example doctrine:

Pragmatist Truth: It is true that *p* if and only if *p* would be believed at the ideal limit of inquiry.[167]

Let's suppose that the pragmatist holds that the preceding proposition, Pragmatist Truth, is a correct description of natural language uses of 'true.' As critics of pragmatism point out, many pragmatists also believe that there will be truths that will not be believed at the limit of inquiry because they remain undecidable, propositions, for instance, about events in the distant past.[168] Suppose this criticism is correct. Then, several pragmatists, C.S. Peirce included, hold inconsistent beliefs and are committed to a conceptual incoherence. It is not possible both that Pragmatist Truth is correct *a priori* and that any true propositions will remain undecidable at the limit of inquiry. Nevertheless, these

[167] See Peirce (1902), p. 565. Peirce gives this now famous characterization saying, "Truth is that concordance of an abstract statement with the ideal limit towards which endless investigations would tend to bring scientific belief."

[168] Peirce (1878) uses the phrase "buried secrets" in anticipating precisely this objection: "But I may be asked what I have to say to all the minute facts of history, forgotten never to be recovered, to the lost books of the ancients, to the buried secrets [â□¦] Do these things not really exist because they are hopelessly beyond the reach of our knowledge?" (207).

conflicting commitments would not disqualify Peirce from political reasonableness, for he would still be a tolerant, rule-following cooperator with operative practical reasoning faculties and would still endorse certain widespread conceptions and values. Indeed, it would seem preposterous to exclude a cooperative, law-abiding philosopher from the political consensus simply because we--or some other authority--had identified a conceptual incoherence in his or her philosophical theory. The surprising upshot is that reasonable people, in Rawls's terminology, can be conceptually incoherent.

Unless a convincing reason can be offered to exclude such persons, that is persons who subscribe to Political Noncognitivism or to Pragmatist Truth or to some other false doctrine, the best course of action for consensus-building is to avoid controversial concepts that would alienate potential members of the overlapping consensus. Thus:

> The advantage of staying within the reasonable is that there can be but one true comprehensive doctrine, though as we have seen, many reasonable ones. Once we accept the fact that reasonable pluralism is a permanent condition of public culture under free institutions, the idea of the reasonable is more suitable as part of the basis of public justification for a constitutional regime than the idea of moral truth.[169]

3. Truth and Consensus

Rawls's desire to circumvent the concept of truth is a symptom of his neutrality towards competing comprehensive doctrines. Rawls articulates this rationale by affirming that No Concept allows political liberalism to incorporate a diversity of reasonable people no matter how their view of political propositions links with truth, whereas Concept Indifference would not have the same breadth for including various comprehensive doctrines within the overlapping consensus.

Where does this leave us? Rawls must choose one, and only one, of the following two options:

1) Maintain No Concept

2) Allow some concept of truth to restrict or determine the set of acceptable political doctrines

Selecting (1) means rejecting Concept Indifference and continuing to eschew the concept of truth, leaving Rawls's project open to the criticisms of Raz, Estlund, and Cohen. Selecting (2) means either (a) adjusting the conception of 'reasonable doctrine' to be more stringent, thereby requiring a less broad consensus, or (b) giving up on the publicity condition and accepting that some reasonable persons will be unable to endorse the

[169] See Rawls (2005), p. 129.

political conception of justice. In other words, Rawls must either accept that his justification of a political conception of justice may be incoherent in virtue of its avoidance of truth, or he must sacrifice the core aim of *Political Liberalism*: providing a justification, endorsed by all reasonable doctrines in a pluralistic democratic society, for a political conception of justice.

In the remainder of this paper, we will defend option (1), maintaining No Concept, as more desirable given the nature of Rawls's project. We will not dispute Raz, Cohen, and Estlund's claims that the justification provided by political constructivism is incoherent in its eschewal of truth. Rather, we will argue that, given the aims of a political conception of justice and the way in which political constructivism seeks to achieve those aims, the question of its coherence or incoherence, when taken alone and unsupplemented, is beside the point. Our key contention on which we base our defense of Rawls and of No Concept is that reasonable people can be incoherent.

IV. Reasonable People can be Incoherent

Does Rawls drop No Concept, or does he allow for the potential incoherence of his constructivist argument for the political conception of justice? Let's evaluate the relative costs of these two options in light of our previous discussion.

1. Dropping No Concept

We have identified two ways to pay the price of incorporating truth into the justification of a political conception of justice. The first is by tightening the conception of reasonableness so that only persons or doctrines that affirm some concept of truth receive full justification from their own point of view. The second is by dropping or weakening the publicity condition, so that not every reasonable citizen can view the governing political conception of justice as acceptable and fully justified.

First, consider Rawls's conception of reasonableness:

> The idea of the reasonable is given in part, again for our purposes, by the two aspects of persons' being reasonable: their willingness to propose and abide by fair terms of social cooperation among equals and their recognition of and willingness to accept the consequences of the burdens of judgment.[170]

The appropriateness of this construal of reasonableness is apparent when one recalls that Rawls's project is to achieve a political conception of justice and a justification thereof that a deeply diverse citizenry can endorse. The conception of reasonableness that Rawls posits is a plausible answer to the question: what are the most basic requirements that citizens must exemplify in order for them to agree on and abide by a single conception of justice? As the stringency of reasonableness increases, the diversity of those who must endorse the governing conception of justice decreases. If we consider reasonable only

[170] See Rawls (2005), p. 94.

those who hold appropriately coherent or true beliefs with respect to normative political statements, then the reasonable constituency to which our justification appeals is smaller, less diverse, and less realistic.[171]

Recall our two concrete cases: Political Noncognitivist and Pragmatist Truth. By hypothesis, both are reasonable in Rawls's sense, yet (we have assumed) both are mistaken or even incoherent when they endorse a conception of justice. The cost of satisfying Raz, Cohen, and Estlund by tightening the conception of reasonableness is the exclusion of such persons from the justificatory constituency. Given the aspirations of *Political Liberalism*, this cost may be prohibitive. Before making this conclusion, however, we must examine Rawls's other options.

The second way in which Rawls could purchase the concept of truth is by weakening the publicity condition. However, this approach is similarly costly. As we have seen, the three levels of publicity all describe ways in which citizens, from their own standpoints, can understand and endorse the institutions and political conception of justice that prevail in their society. At various points, Rawls gives detailed reasons why publicity is an important desiderata for a just society.[172] Ignoring the details of Rawls's reasons, it remains apparent that without satisfying one of the three levels of publicity, there will be some subset of reasonable persons who cannot endorse the prevailing political conception of justice. Either they cannot accurately affirm that society and its institutions satisfies a conception of justice they endorse (first level), they do not agree with the empirical facts or methods of inquiry that support the justification of the prevailing conception of justice (second level), or they cannot know or do not have access to the argument used to justify the prevailing conception of justice (third level). In all cases, a set of reasonable persons is unable to endorse the political conception of justice and Rawls fails to achieve the stated aim of *Political Liberalism*.

The cost of satisfying Raz, Cohen, and Estlund and of avoiding incoherence thus appears to be quite high. What is the cost of continuing to eschew truth at the risk of presenting an incoherent justification?

2. Eschewing Truth

Raz, Cohen, and Estlund's basic strategy is to find a contradiction in the approach taken in *Political Liberalism* by claiming that Rawls smuggles in or depends upon the concept of truth all the while claiming to avoid it. Rather than examining each position separately, as we will in the next section, consider a general formulation of the objection in terms of the norm of assertion. It is held by many that knowledge is the norm of assertion.[173] That is to say, when someone asserts that *p*, they are appropriately subject to blame unless they know that *p*. Now political discourse inevitably involves numerous assertions. Parties

[171] See Huemer 1993.
[172] Three such reasons, identified by Kogelmann 2017, are public scrutiny, social unity and autonomy.
[173] See Williamson (2000), ch. 11.

engaged in such discourse claim rights and corresponding obligations, they make evaluations to bolster these claims, they assert the appropriateness of certain economic and moral tradeoffs. Such assertions certainly characterize Rawls's approach of political constructivism. If successful assertion requires knowledge, then truth is also implicated, for it is universally accepted that knowing that *p* entails the truth of *p*. Thus, if knowledge is the norm of assertion, then involving assertion in political deliberation violates No Concept. Assuming this line of reasoning is correct, the cost of eschewing truth in political constructivism is coherence. A justification of a conception of justice is nonsensical without recourse to the concept of truth.

The defender of Rawls might think that, in order to avoid this difficulty, the burden is to show that knowledge is not the norm of assertion. Perhaps one could find another norm of assertion that is not conceptually linked to truth,[174] and argue that this is actually the correct norm of assertion. However, such an approach misses the point. Given the aims of *Political Liberalism*, political constructivism is unconcerned with what the norm of assertion actually is. It need not take a stand on this philosophical question.

Suppose that it is true that knowledge is the norm of assertion. Presumably a person could be reasonable according to Rawls's definition yet nevertheless disagree that knowledge constitutes such a norm. Because of this reasonable disagreement, political constructivism--aiming to attain society-wide consensus--would, nevertheless, take no stand on what assertion entails, instead leaving that up to the determination of each comprehensive doctrines. But since knowledge actually is the norm of assertion, large swaths of reasonable people will be committed to something false. This is unsurprising; there is no requirement that reasonable people only believe what is true in their comprehensive doctrines.

What is surprising is how far this ignorance can extend. Let us define conceptual incoherence as a state where one's beliefs about a particular concept are ultimately contradictory. This would occur if a person were to believe that knowledge is the norm of assertion and had beliefs that contradicted that outright or via some entailment, perhaps by believing that knowledge is the norm of assertion, that knowledge entails truth, and that the norm of assertion has no connection to truth. Conceptual incoherence is obviously an undesirable state. When one involves themselves in conceptual incoherence, this means that at least one of their beliefs is false, and it is not clear that one can be both rational and conceptually incoherent. Reasonable people, however, can be conceptually incoherent. The inconsistency of an individual's or of a doctrine's beliefs with respect to a topic such as the nature of normative-political truth, of the norm of assertion has little if any bearing on their willingness to cooperate according to fair terms and to recognize the burdens of judgment. A diverse society can therefore embrace such individuals and doctrines under a conception of political justice.

In response to Raz, Cohen, and Estlund, therefore, the defender of Rawls can simply say

[174] An example would be Rachel McKinnon's (2015) sufficient reason norm of assertion (p. 4).

that the cost of justificatory incoherence is low. Some comprehensive doctrines may espouse conceptually coherent and factually accurate beliefs, they may supplement the bare-bones justification offered by Rawls with a theory of normative truth.[175] But participants in the overlapping consensus can also hold positions that are incorrect, even on pain of conceptual incoherence, without ceasing to be reasonable. The aim of *Political Liberalism* is to reconcile these disparate doctrines, to find a conception that a diverse citizenry "religious and nonreligious, liberal and nonliberal" --coherent and incoherent, we add--"may endorse for the right reasons."[176]

V. A Final Response to the Truthers

With this understanding in place, what should we make of Raz, Cohen, and Estlund's particular criticisms? We have already shown that the charge of incoherence can be assuaged by the simple fact that parties to the overlapping consensus can be conceptually incoherent. To drive the point home, we examine each critic's view in greater detail.

1. Raz

Returning to our discussion on the norms of assertion, a Rawlsian response to Raz becomes immediately available. Recall that Raz's worry was that recommending a principle of justice entails regarding that principle as true. While the knowledge norm of assertion implicates truth as conceptually connected to assertion, Rawls may respond that the proper norm of assertion in political discourse should be that an assertion is acceptable to a reasonable citizenry. Given the task of a political conception of justice, the debate is not over what the norm of assertion actually is, but what notion of assertion all parties in an overlapping consensus could agree on. The aim of a certain kind of discourse should have a bearing on the norms that govern it. In political discourse, especially of the justificatory type, the goal is not to describe a mind-independent reality.[177] Instead, the task of political discourse – for us "here and now" – is related to solving the problem of political liberalism, viz., "to work out a political conception of political justice for a constitutional democratic regime that a plurality of reasonable doctrines, both religious and nonreligious, liberal and nonliberal, may endorse for the right reasons."[178]

If political discourse has this practical task, then the norms that govern political discourse must not generate unnecessary faction or controversy. Different reasonable doctrines have radically different positions regarding the status and relevance of truth. Thus, a more sensible norm would be one that all parties can endorse as in conformity with the values and concerns that motivate political discourse. Reasonableness, understanding a "reasonable assertion" to be one which all involved parties can accept insofar as they are

[175] See PL 144-5.
[176] Ibid, p. xxxix.
[177] See Rawls (2005), pp. 91-93
[178] Ibid, p. xxxix.

reasonable, is (by definition) the norm that fulfills this requirement. Because the reasonableness norm is premised on appealing to the reasoning faculties of all involved parties, it fosters consensus rather than discord, thereby fulfilling the task of political discourse. Therefore, even if we accept Raz's claims about normative-political truth, it does not follow that political constructivism should seek to avoid the incoherence that he has identified. Political constructivism provides an argument that may be supplemented in coherent or incoherent ways by diverse comprehensive doctrines. Whether it is coherent or incoherent taken alone without being backed by any theory of truth is not relevant, because the argument is not meant to be taken alone. It is, instead, a "module… that in different ways fits into and can be supported by various reasonable comprehensive doctrines that endure in the society regulated by it."[179]

2. Cohen

The case of the Political Noncognitivist should not mislead us into thinking that Rawls is endorsing a noncognitivist understanding of political assertions. This is important to how Rawls would respond to Cohen's criticism. Cohen argues that by using concepts that are conceptually connected to truth, Rawls commits his political constructivism to something beyond No Concept. As Cohen puts it, "Truth is so closely connected with intuitive notions of thinking, asserting, believing, judging, and reasoning that it is difficult to know what leaving it behind amounts to."[180] As in the Rawlsian response to Raz sketched above, the issue is not that Cohen is wrong about the nature of truth or of assertion. Rather, it is that many reasonable people disagree, and therefore a norm of assertion that involves truth instead of mere reasonableness is unduly exclusive. One class of such people is noncognitivists, who hold that normative statements are not truth-apt. If a noncognitivist were persuaded by Rawls's constructivist argument to endorse the political conception of justice and to abide by its demands, then whether they think the principles of justice are "true" or not is beside the point. This is why Rawls aims to construct the conception out of materials that are drawn from a public culture, rather than from some particular view of what is morally worthy or true.

Rather than putting forward reasons for his principles of justice on the basis of their truth, Rawls proposes a procedure of construction by which each citizen can see the principles as issuing from their own practical reason and normative conceptions. Doing so does not require that Rawls commit himself to any theory of the truth-aptness of normative claims. In his critique, Cohen exhibits a serious confusion on this point by arguing that Rawls is committed to a cognitivist view of normative political statements. First Cohen argues that:

> The claims made by a political conception... must be truth-apt ... They must be, if there is to be a common ground of argument under conditions of doctrinal disagreement. To deny the truth-aptness of the claims made on the terrain of public reason would offend against the essential idea of public reason. That is

[179] *Political Liberalism*, 145.
[180] See Cohen (2009), p. 15.

> because the very propositions advanced in public political argument, even if not taken as or presented in that context as true, might be judged to be true by the religious or moral doctrine affirmed by a citizen.[181]

The truth in what Cohen says is that Rawls is not free to deny the truth-aptness of normative claims. Doing so would alienate moral or religious doctrines that judge such claims to be true or false. But from this, Cohen goes on to infer that "Rawls's proposal is to endorse a cognitivist understanding of political conceptions of justice and political argument on which notions of judgment, reasoning, and argument are fully in play, while denying the availability of the concept of truth within such conceptions."[182] This inference, however, is clearly mistaken. It would be just as illicit for Rawls to endorse and argue from a cognitivist view as a noncognitivist view. In either case, there are reasonable comprehensive doctrines, affirmed by reasonable citizens, that reject the metaethical view in question. Accordingly, the proper path for political constructivism is to avoid taking a stand on this metaethical question.[183] Doing so would undermine the project of *Political Liberalism* and thwart the task of political constructivism by rendering the procedure unpersuasive, unacceptable, or even incomprehensible, to a large group of reasonable citizens.

3. Estlund

Finally, the constructivist argument that we sketched above also brings into focus the Rawlsian response to Estlund's critique. Estlund points out that acceptance by a particular group is insufficient for a doctrine to gain normative import. After all, there are many such "insular" groups, like the Star Davidians. In order to justify privileging one normative criteria over others, one must hold that it is true, not merely reasonable.

The Rawlsian response should be clear by now: we do have good reason to favor reasonableness over Star-Davidianism as giving rise to obligations, but this reason is not its truth. Rather, the standard of reasonableness is tightly connected to that of acceptability to the relevant constituency, i.e. those willing to cooperate, i.e. reasonable persons.[184] A conception of justice is reasonable to the extent that it can gain acceptance among reasonable persons in reflective equilibrium.[185] That is, a reasonable conception "meshes with and articulates our more firm considered convictions,"[186] and in this way seeks justification within the constraints of each individual's own practical reason and set of values. If the aim of a political conception of justice is to make mutually acceptable

[181] Ibid, p. 18.
[182] Ibid, p. 19.
[183] As Rawls (1999) says, "It is important to notice here that nor assumptions have been made about a theory of truth. A constructivist view does not require an idealist or a verificationist, as opposed to a realist, account of truth." (p. 351).
[184] More specifically, the conception of reasonableness and its normative import emerge from the shared materials of construction: practical reason, the social role of justice, and our publicly shared conception of person and society.
[185] See Rawls (1999), p. 321.
[186] Ibid, p. 321.

the institutions and claims to which each of us is subject, then reasonableness is clearly special in its ability to facilitate that aim. In fact, it is definitionally true that the reasonable doctrine is the one most capable of justifying the basic structure to each citizen. Thus, it is the political task that we are engaged in and the moral bases that such a task implicates that favor the criteria of reasonableness over that of Star-Davidianism.

Works Cited

Cohen, Joshua. 2009. "Truth and Public Reason." *Philosophy and Public Affairs* 37, no. 1: pp. 2-42.

Estlund, David. 1998. "The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth." *Ethics* 108, no. 2: pp. 252-275.

Kogelmann, Brian. "Justice, Diversity, and the Well-Ordered Society," Forthcoming in *The Philosophical Quarterly*, 2017.

McKinnon, Rachel. 2015 *The Norms of Assertion: Truth Lies and Warrant.* Palgrave Macmillan.

Peirce, C.S. 1878. "How to Make Our Ideas Clear." In *The Nature of Truth: Classic and Contemporary Persectives. Ed. Michael Lynch. MIT Press, 2001: pp. 193-210*

Peirce, C.S. 1902. "Truth and Falsity and Error." *The Collected Paper of Charles Sanders Peirce*, vol. 5, ed. Charles Hartshorne and Paul Weiss. Harvard University Press, 1965: pp. 565-573.

Rawls, John and Freeman, Samuel. 1999. *John Rawls: Collected Papers*. Harvard University Press.

Rawls, John. 2005. *Political liberalism*. Columbia University Press.

Raz, Joseph. 1990. "Facing diversity: The Case of Epistemic Abstinence." *Philosophy and Public Affairs* 19, no.1: pp. 3-46.

Williams, Bernard. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton University Press.

Williamson, Timothy. 2002. *Knowledge and its Limits*. Oxford University Press.

Wenar, Leif. 1995. "Political Liberalism: An Internal Critique." *Ethics* 106, no. 1: pp. 32-62

Answerability Without Blame?*

Andrea Westlund

Abstract: Though widely derided by popular psychologists and self-help writers as an emotionally toxic and destructive response, blame has many defenders among contemporary moral philosophers. I argue that some disagreement over the value of blame can be explained by the fact that blaming, as a speech act, takes several distinct forms. Popular critiques of blame, I suggest, properly target what we might call judgmental or strongly verdictive blaming, the sort of blaming that passes judgment on the wrongdoer him- or herself and treats him or her as deserving of the blamer's hostile or "punishing" reactions. Such reactions tend to foreclose further moral dialogue with wrongdoers. In many contexts, I argue, it is more appropriate – and more constructive – to hold others answerable *without* blaming them in the strongly verdictive sense.

Blame has many defenders these days. George Sher treats blame as inseparable from morality and as called for in response to the violation of moral principles we endorse (Sher 2007). R. Jay Wallace takes blame to be a form of "deep moral assessment" (Wallace 2008, 179) that is intimately connected to the practice of holding people to moral expectations. Macalester Bell, similarly, takes blame to be crucial to our responsibility practices, and considers the standing to blame to be inalienable – it is of such central importance to our status as valuers, she argues, that we cannot be stripped of it in virtue of hypocrisy, complicity, or other such failings. Blame is not the preserve of the morally pure but an important exercise of moral agency for all valuers (Bell 2012).

A quick survey of the titles of popular self-help books about blame tells a different story. Take, for example, the particularly colorful *Beyond Blame: Freeing Yourself from the Most Toxic Emotional Bullsh*t* (Alasko 2011). This book, like others in its genre, treats blame as an entirely destructive emotional phenomenon. According to its author, to blame is to *find fault* with others (using criticism, accusation, punishment, or humiliation) and to *shift responsibility* for one's own behaviors on to others, in order to avoid being seen as wrong or bad oneself. The latter tactic is presented as a defensive maneuver, and it is not hard to see why it is supposed to be toxic: insofar as blaming someone else is just a sneaky way of avoiding accountability for oneself, there's not much to be said in its favor. It is not clear, however, why we should suppose that blame

* *This paper is a portion of a longer paper, in which I go on to discuss additional cases and examples and further develop the ideas toward which I gesture in the concluding paragraph. I hope, however, that these first two sections stand well enough on their own for purposes of a conference presentation.*

is always or even typically misdirected in this way – and if this presupposition does not hold up, the reach of the responsibility-shifting critique will be limited.[187]

The critique of "fault-finding", by contrast, does not rely on assumptions about the (mis-) assignment of responsibility. Carl Alasko, in the above-cited text, claims that the hostile and accusatory attitudes associated with blame are always destructive, and that blame damages relationships by representing other persons as flawed or defective in virtue of the fault. On these points, the concerns of self-help writers like Alasko are shared by at least some philosophers.[188] While I do not myself endorse a thorough-going rejection of blame, concerns about the destructiveness of fault-finding nonetheless strike me as containing a kernel of insight.[189] I want to push against the grain of recent defenses of blame just far enough to articulate what this popular line of critique gets right.

In this paper, I distinguish between blame as a reactive attitude and blam*ing* as a speech act, and argue that some disagreement over the value of blame can be explained by the fact that blaming, as a speech act, takes several distinct forms. The critique of fault-finding, I suggest, properly targets what we might call judgmental or strongly verdictive blaming, the sort of blaming that passes judgment on the wrongdoer him- or herself, and treats him or her as deserving of the blamer's hostile or "punishing" reactions. This kind of blaming, I argue, tends to foreclose engagement in further moral dialogue with wrongdoers, and such dialogue is something that we ought to care about for a variety of reasons – not least because of the way in which it underwrites and supports agents' capacity to hold themselves and others answerable.

The problems with strongly verdictive blaming are perhaps most visible in therapeutic or pedagogical contexts, where passing judgment on a wrongdoer runs counter to the aim of encouraging her to hold herself answerable and take responsibility for her actions. I therefore begin, in §1, by considering Hannah Pickard's recent account of responsibility without blame in therapeutic settings. In §2 I build on her account by distinguishing between blame (the attitude) and blaming (the speech act), and explain what I mean by "strongly verdictive" blaming. I argue that strongly verdictive blaming clearly runs counter to the therapeutic aims identified by Pickard, but that we can hold others answerable, and even be angry with others, without blaming them in this sense. I conclude by challenging the idea (taken for granted by some neo-Strawsonians) that there is a sharp dividing line between therapeutic and non-therapeutic responses, or between

[187] The responsibility-shifting critique rests on the background presumption that blamers are rarely (if ever) *justified* in assigning responsibility to the others – a presumption that is clearly questionable, and potentially pernicious, since it may serve to silence or dismiss the expression of legitimate concerns and claims by those who have been wronged.

[188] Martha Nussbaum, for example, has recently argued that anger as a response to wrongdoing is nearly always irrational and deeply destructive, and recommends replacing it with forward-looking attitudes of unconditional love and generosity (Nussbaum 2016). Glen Pettigrove is similarly suspicious of emotions in the anger family, and argues that "meek" responses to wrongdoing have significant epistemic and moral advantages (Pettigrove, 2012). Although these arguments are focused on negative emotions associated with blame, rather than on blame itself, they clearly share the view that harshly critical, accusatory, or vilifying reactions to wrongdoers are ethically flawed and do more harm than good.

[189] I will say more about the relationship between anger and blame below. I explore reasons for disagreeing with Nussbaum's assessment of anger in my …

objective and participant-reactive attitudes, toward wrongdoers. "Holding answerable", I suggest, is a dialogical response that may have a therapeutic or pedagogical dimension without thereby objectifying its target.

§1

In her paper "Responsibility Without Blame" (2011), Hannah Pickard argues that in certain clinical contexts it is crucial that service providers treat service users as responsible without subjecting them to blame. Pickard's focus is on the case of effective treatment of personality disorder (henceforth, PD). Service users with PD are responsible, she argues, in the sense that they are generally consciously aware of what they are doing and exercise choice in doing it. In other words, they meet basic epistemic and control conditions for responsible agency. Responding to these service users as if they were *not* responsible – treating, them for example, as passive victims of their troubled personal histories – is counterproductive with respect to therapeutic goals, which include the goals of getting the service users to *take* responsibility for their behaviors and to choose to act differently, and more constructively, in the future. But responding to these service users with blaming attitudes and behaviors is also known to be disruptive of these therapeutic goals. Pickard takes on the challenge of articulating a stance that avoids both pitfalls, rescue and blame, a stance to which she refers as "responsibility without blame".

What accounts for blame's interference with therapeutic goals, according to Pickard, is its characteristic "sting" – "[e]ffective treatment," she observes, "is not possible if the service user feels judged, shamed, berated, attacked, or hurt" (Pickard 216). Pickard describes blame as a "punishing" mental state (2011, 219), which may be expressed through actual punishment but which "can also be manifest in berating, attacking, humiliating, writing off, rejecting, shunning, abandoning, and criticizing" (2011, 218), among other things. There is a striking similarity between this list of behaviors and the that put forth by Alasko (2011) in the critique of fault-finding mentioned above. But Pickard makes a further, insightful point about the structure of blame, namely that it is not just a collection of negative attitudes and dispositions to behave, but a mental state in which these attitudes and dispositions are united by the feeling that one is *entitled* to feel them toward the offender, in response to the offenders' behavior, and that one's hostile emotional reactions are *deserved* by the wrongdoer.

In recommending responsibility without blame, Pickard is in effect recommending a kind of blame that is stripped of this sense of entitlement – or, at least, a kind of blame in which the tendency to act on one's sense of entitlement has been tamped down. The model Pickard has in mind is the stance of a service provider who judges a service user to be responsible – and, indeed, *morally* responsible and thus blameworthy – for problematic behavior, without actually blaming them in the hurtful, affective sense. Pickard argues that providers who take this stance *hold* users responsible (indeed, hold them blameworthy) without having, as she puts it, "a feeling of entitlement to any negative reactive attitudes and emotions one might experience, no matter what the service user has done" (Pickard 2011, 219). Their blame is detached: it includes a *judgment* of blameworthiness, but no corresponding *feeling* of entitlement to the attitudes and

emotions characteristic of affective blame.[190] Without that feeling of entitlement guiding one's reactions, Pickard argues, one can allow compassionate consideration of the service user's troubled background to moderate one's anger or resentment, keeping it in check while one holds the user accountable in more constructive ways.

Pickard paints a compelling picture of what it might be to hold responsible without what she calls *affective* blame. There is, however, a puzzle at the heart of this picture. "We judge a person to be blameworthy," Pickard tells us, "when they are responsible for harm, and have no excuse" (2011, 215). This tells us the conditions under which we rightly judge a person to be worthy of blame, but not yet what it is that such a person is judged to be worthy *of*. What exactly is it, of which one is judged worthy, when they meet these conditions? Surely it must, on pain of vicious circularity, be the attitudes and emotions involved in affective blame. But if one judges that an offender is worthy of affective blame (in other words, that affective blame would be "fitting" with respect to such an offender), then the feeling of entitlement that is part and parcel of affective blame would be fitting as well – one would be every bit as entitled to one's anger and resentment as one felt. And, if that feeling of entitlement would be fitting, it appears that it would take a bit of self-hoodwinkery to convince oneself one ought not to feel it.

In the case of PD, there are, of course, reasons for not allowing oneself to feel the sense of entitlement characteristic of affective blame, or at least not allowing oneself to express or act on the various negative emotions to which one feels entitled: as we've already seen, it would be better, from the point of view of clinicians' therapeutic aims, *not* to feel it (or at least not to express it), because then one is able to engage the service user more effectively in interactions that will encourage her to take responsibility for problematic behavior and to refrain from such behavior in the future.[191] But are these instrumental reasons the *right kind of reasons* for not feeling or expressing affective blame? Do we not, in virtue of treating the user compassionately for therapeutic reasons, slip out of the participant-reactive stance altogether, and into what Strawson calls the objective stance, and thus in the end fail to treat them as responsible after all? The power of the example of PD resides in the convincingness of the claim that individuals with PD *are responsible*, so it would be a theoretical loss to have to retreat to a stance on which we merely pretend to react to them as such (engaging in "as if" behavior), while in fact stepping outside the realm of genuinely participant-reactive attitudes.

I think Pickard is right that service providers can in principle (and properly speaking) hold service users responsible without blaming them in the problematic, hurtful sense. They are not simply treating service users as forces of nature, to be managed rather than reasoned with. But the appeal to detached blame – which, if I am right, is just an appraisal concerning the fittingness of affective blame – does not give us all the tools we need to understand their stance. In the next section I suggest that we must distinguish not only between affective and detached blame, but also between blame as a reactive

[190] *Cold* blame, as Wallace and others construe it, seems to me compatible with feeling entitled to the negative emotions involved in affective blame, but just not feeling those emotions themselves. So I will take it to be a slightly different attitude from the one Pickard describes as detached blame.

[191] Pickard discusses various techniques that service users might employ – including focusing one's attention on the service user's past history – to tamp down counter-productive impulses.

attitude and blaming as an expressive act. We will then be in position to consider various forms that blaming may take, and the aspects of our responsibility practices that are supported (or undermined) by these forms.

§2

Blame, understood as a reactive attitude, is closely associated with the speech act of *blaming* – an act that is sometimes explicitly performed through utterances such as "I blame you", but sometimes instead implicitly expressed through acts of criticism, censure, and the like. I propose that focusing on the uses of blame, instead of on the attitudes felt by blamers, may help us move forward. Blame in the speech-act sense makes several (brief) appearances in J. L. Austin's classic text *How to Do Things With Words*. Most prominently, it is offered as an example of a behabitive – or in other words, of a performative that exhibits, as opposed to merely describing, attitudes and feelings (Austin 1962, 83).[192] In classifying blame as a behabitive, Austin places it in what he describes as "a very miscellaneous group, … [having] to do with attitudes and *social behaviour*" (Austin 1962, 88, emphasis in original). But Austin later notes that blame also has verdictive and exercitive uses. What do these add to the behabitive use?

Verdictives, according to Austin, involve the giving of a verdict or the delivering of a finding with respect to matters that require the exercise of judgement. In official contexts, verdictives are delivered by judges or others in positions of authority, and many of Austin's examples are drawn from the realm of law. But he also suggests that informal blaming has a verdictive sense, which he takes to be equivalent to holding responsible (Austin 1962, 155).[193]

Exercitives, by contrast, involve the exercise of "powers, rights, or influence" (Austin 1962, 151). I take this to mean that they involve the exercise of what many philosophers now refer to as normative or moral powers – powers to make moral claims, give reasons, impose obligations, or otherwise alter the normative landscape just through expressing an intention to do so. (Austin's initial list of examples clearly fall in this category: "appointing, voting, ordering, urging, advising, &c" (151).) According to Austin, an exercitive "is a sentence rather than a verdict" (151). He does not explicitly argue that blame has an exercitive use, but he clearly implies that it does when he cites "*non*-exercitive uses of blame" as an example of a behabitive.

When considering blame as a speech act, it is important to notice that the term as we commonly use it is ambiguous between its verdictive, exercitive, and behabitive senses. When one takes into consideration the possibility that these three uses may in practice overlap and be combined in various different ways, the prevalence of sharp disagreement over the value and significance of blame begins to seem much less surprising (and, indeed, much less intractable). Defenders and critics may well be focusing on different senses, and different uses, of blame.

---

[192] Strictly speaking, Austin treats blame as "half descriptive" rather than purely performative. Blame's purely performative counterparts, in the passage under discussion, are criticism and censure (Austin 1962, 83).

[193] I would quibble with this equivalence, since I think "holding", especially as it figures in the phrases "holding responsible" and "holding answerable", itself has non-verdictive senses. I use these phrases in a non-verdictive sense myself in what follows.

Let us return to the case of PD, and ask ourselves what it is to which service users are responding when they are blamed in a way that makes them feel "judged, shamed, berated, attacked, or hurt" (Pickard 2011, 216). The fact that they feel *judged* strongly suggests that they are reacting to blame in its verdictive sense – a verdict has been passed, and it is not a favorable one. The fact that they at the same time feel shamed, berated, attacked, and hurt, suggests that the verdict has been accompanied by the targeted expression of "punishing" attitudes, as if an implicit sentence has been carried out through the expression of blame itself. One experiencing these reactions experiences herself as (metaphorically) having been convicted, sentenced, and punished all at once: the blamer acts as judge, jury, and executioner. It is not hard to see why blame in this sense – blame that is at once verdictive, exercitive, and behabitive – is contrary to therapeutic goals. This blame – to which I'll refer as "*strongly* verdictive" – is judgmental and punitive, and implicitly puts the blamer in a position of authority over the one blamed. One might, of course, contest such blame, but it is not the sort of thing that naturally draws one into moral dialogue. In the face of such an onslaught, it may be less emotionally costly to disengage than to defend oneself.

Let's consider, by contrast, how service providers (as described by Pickard) regard and respond to service users, and how service users with PD respond to the overtures of their service providers, when all goes well. We are given the impression that in such cases, service providers hold service users responsible, and service users respond by coming to *hold themselves* (increasingly) responsible for their actions, a stance that allows them to make different and better choices in the future. Such an outcome must, presumably, proceed from a shared sense that current or past behavior has been problematic. But the passing of a verdict seems not to be central to the point of this interaction. (Certainly, the point is not to pass a verdict on the *person* as defective or faulty in light of their behavior, which is the kind of maneuver that self-help writers find so destructive.) Service users begin from a position of imperfect self-responsibility, but insofar as they are responsive to therapeutic interventions, they manifest the capacity to hold themselves responsible with the right sorts of promptings from others – and service providers interact constructively with them by proceeding on the assumption that this capacity is indeed in place, and not irredeemably faulty.

Agents with PD thus appear to lie at the margins (but, importantly, not *outside* the margins) of our responsibility practices, insofar as these practices presuppose the capacity to hold oneself and others responsible. The point of engaging service users in the way that service providers do, is to draw them in, from the margins, somewhere closer to the center. Reacting with the trifecta of verdictive, exercitive, and behabitive blame does not support these agents' incipient responsibility competence, but *holding them answerable* does. So what is it to hold someone answerable, and how is it different from (or related to) *blaming* them?

Holding answerable, like blaming, is both an attitude and an act. In the act sense, holding another answerable seems clearly to be exercitive: it involves exercising a moral power to demand a response from another, and puts them under a (defeasible) obligation to respond. As such, holding answerable may seem to presuppose something like verdictive blame, at least in the sense of a *pro tanto* finding of wrong-doing. This seems right, but it is important to note that holding answerable is nonetheless not *strongly*

verdictive in the sense outlined above. As an exercitive it is akin not to a sentencing, but to a summons to moral dialogue. The prospect of dialogue highlights the *pro tanto* nature the accompanying verdict, and makes it defeasible in practice and not merely in principle. The meaning or significance of an action is typically subject to competing interpretations, particularly in morally complex situations. Currents of answerability run in more than one direction, and may lead to mutual moral insight and growth – as well as, in some cases, the sharing of associated moral burdens or costs. Holding another answerable, in other words, is not just a way of checking whether a person is indeed blameworthy, by scanning her answers for considerations that might excuse or justify. In holding another answerable, one invites her into an alternative perspective on her action, and opens up the possibility of coming to a constructive shared response.

The dialogical structure of holding answerable in the exercitive sense also casts any accompanying behabitives in a different interpretive frame: even the exhibition of negative emotions such as anger may be read as demanding and engaging rather than as punitive and vilifying (more on this below). In holding another answerable, one implicitly treats her as a moral peer, or at least as one who *can* be a moral peer, rather than shaming, humiliating, or otherwise down-grading her in virtue of a moral fault.

This brings us to the attitudinal side of holding answerable. To hold a person answerable is to take a practical, affective stance toward them, and indeed, seems to count as having a reactive attitude. Lucy Allais helpfully defines a reactive attitude as "an affective way of regarding a person, which involves being disposed to have a range of feelings toward her in a range of circumstances. It involves seeing her in a certain way, being disposed to have characteristic patterns of attention, interpretation and expectation with respect to her actions" (Allais 2008, 7). It is not hard to see that treating someone as answerable involves certain patterns of attention, interpretation and expectation with respect both to her dialogical responsiveness and to her future decision-making and action. One will attend to her responsiveness or lack thereof to one's challenges and suggestions, will interpret her replies or evasions as the replies or evasions of one who can be held to relevant normative standards (not as a force of nature), and will be disposed to hold her to expectations regarding how she carries the result of such dialogue over into her future decision-making and action.[194]

It might be more difficult to establish what feelings are central to the attitude of holding answerable. I would argue that while anger is not *required* for holding answerable, it has a proper place here: I take anger to be an affective appraisal of wrongdoing that demands (or is such as to demand) a response from the wrongdoer. Those who take anger always to be destructive tend to argue that it conceptually includes a desire for payback (Nussbaum 2016) or a desire to lash out (Pettigrove 2012). I don't think these punitive desires are essentially connected to anger, but I will not try to settle that issue here. Whether or not one takes anger to have a proper place in holding answerable, one will feel other emotions – disappointment, gratification, frustration, or

[194] Wallace (2008) argues that holding another or oneself to a moral expectation must itself be understood in terms of dispositions to respond with blame (and corresponding attitudes such as guilt). On my view holding to an expectation does require dispositions to respond in a way that marks unwillingness simply to "let go" of the behavior in question, but I think there are ways of refusing to let go that do not involve the anger or resentment characteristic of blame.

hope, for example – depending on how one's interactions with the agent go, and one may certainly experience friction or harmony with the other and be motivated to change tack or tone as appropriate in light of such experiences. Holding answerable as a practical stance, or reactive attitude, will indeed include disposition to engage in corresponding speech acts of holding answerable – the attitude and the act, though not inseparable, tend to go hand in hand in our everyday responsibility practices.

In centering the attitudes and speech acts involved in holding answerable, my reading of interactions with agents with PD supports a variation on a communicative approach to responsibility. Defenders of communicative accounts of responsibility argue that meeting the control condition on responsible agency depends in part on the responsible agent's having the capacity to hold herself and others responsible. On Michael McKenna's view, for example, the morally responsible agent is analogous to a competent speaker of a natural language – a speaker, that is, who is conversationally adept in that language. Much as a competent speaker's speaking skills are enmeshed with her interpretive skills (her ability to interpret and understand what other speakers of the language are saying, and to allow relevant norms of meaning to guide her own conversational forays), a morally responsible agent's competence as a responsible actor is enmeshed with her competence as an interpreter of the actions of others. One's ability to appreciate the significance of what others do, insofar as it manifests the quality of their will, is entwined with one's own ability coherently to manifest the quality of one's own will in one's own actions. One who is at sea in this system of "agent meaning" is not fully morally responsible, in a manner analogous to that in which a speaker who is at sea with respect to a language she cannot competently speak is not fully linguistically responsible for what she says or fails to say. On McKenna's model, an act performed by a morally responsible agent is analogous to an opening gambit in a conversation – a gambit to which others respond with further moves, which themselves invite yet further responses. Competent moral agents know their way around complex practices of praising, blaming, excusing, repenting, apologizing, forgiving, and so on and so forth.

What the PD cases suggest is that it is of even more fundamental importance that competent moral agents know their way around practices of holding and being held answerable – or, at least, that they are capable of finding their way around these practices with the right sort of support, guidance, or prompting from others. Agents who "shut down" in the face of strongly verdictive blame may still be responsible, and appropriately *held* responsible, insofar as they are competent (enough) participants in our answerability practices.[195]

One thing that is brought out nicely by the example of PD, is that the idea that there is a clear and sharp line between those who are capable of holding themselves and others responsible and those that are not is something of an artifice.[196] There are, to be

[195] McKenna does make mention of the fact that our responsibility practices include practices of holding answerable, but they are not treated as central to the ways in which we hold others answerable for morally criticizable acts (to which he refers as "blameworthy" acts). Holding others answerable is said to be part of what it is to hold someone responsible in the status sense, but when it comes to particular acts, answerability practices seem to be treated more in the manner suggested in section 1 above, as information-gathering preludes to blame and its correlates.

[196] Strawson himself acknowledges that a "simple opposition" between the participant-reactive and objective stances is too crude, and observes that there will be cases that straddle the two kinds of attitude.

sure, those who cannot be drawn into our responsibility practices, just as there are those who cannot be drawn into linguistic practices, because they lack underlying cognitive and (perhaps) emotional capacities. And perhaps (though I find it less likely) there are those who are perfectly fluent. But then there are the rest of us: the large majority of people who exhibit varying degrees of fluency or disfluency in both kinds of practice, in ways that fluctuate across the span of a life, and vary across different local sub-practices or "idiolects". In morally engaging one another, imperfectly responsible agents attempt to elicit, from one another and themselves, *more* perfectly responsible behavior. Individuals with PD may indeed be marginal, as far as their competence within the relevant practices goes, but they are not unreachable, and the form that therapeutic intervention takes with such individuals bears out this important point. Responsible agency is always to some degree a work in progress, and responses that strike us as therapeutic, insofar as they contribute to such progress, may nonetheless be fully participatory, insofar as the operate by engaging others *as practitioners* (at least in prospect) from within the practices in question.

References

Alasko, Carl. 2011. Beyond Blame: Freeing Yourself from the Most Toxic Form of Emotional Bullsh*t. New York: Tarcher/Penguin.

Allais, Lucy. 2008. "Dissolving Reactive Attitudes: Forgiving and Understanding," South African Journal of Philosophy. 27(3): 1-23.

Austin, J. L. 1962. How to Do Things With Words. 2nd Edition. Ed. J. O. Urmson and Marina Sbisa. Cambridge, MA: Harvard University Press.

McGeer, Victoria. 2012. "Co-reactive Attitudes and the Making of Moral Community," Emotions, Imagination, and Moral Reasoning. Ed. C. MacKenzie & R. Langdon. New York: Psychology Press.

McKenna, Michael. 2012. Conversation and Responsibility. Oxford and New York: Oxford University Press.

Nussbaum, Martha. 2016. Anger and Forgiveness: Resentment, Generosity, Justice. Oxford: Oxford University Press, 2016.

Pettigrove, Glen. 2012. "Meekness and 'Moral' Anger," Ethics 122: 341-370.

Pickard, Hannah. 2011. "Responsibility Without Blame: Empathy and the Effective Treatment of Personality Disorder", Philosophy, Psychiatry, & Psychology. 13(3): 209-224.

He discusses parents' attitudes toward children in this regard, as well as therapists' attitudes toward their patients. In the case of children, he describes parents' attitudes as a compromise, shifting back and forth "between objectivity of attitude and developed human attitudes" (Strawson 1974, 32). In the case of the therapist, he says "*His* objectivity of attitude, *his* suspension of the ordinary moral reactive attitudes, is profoundly modified by the fact that the aim of the enterprise is to make such suspension unnecessary or less necessary" (Strawson 1974, 32). What Strawson describes in these passages still seems to involve more of a *suspension* of participant-reactivity than I have in mind. I will argue that some modes of participant-reactivity themselves have a therapeutic or quasi-therapeutic dimension, such that one is not shifting back and forth between attitudes but adopting an attitude that combines features of both.

[1]

Scanlon, T. M. 2008. Moral Dimensions: Responsibility, Meaning, Blame. Cambridge, MA: Belknap Press of Harvard University Press.

Sher, George. 2007. In Praise of Blame. Oxford and New York: Oxford University Press.

Shoemaker, David. 2011. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," Ethics. 121: 602-632.

Smith, Angela. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life," Ethics 113: 236-271.

Strawson, P. F. 1974. "Freedom and Resentment," reprinted in Freedom and Resentment and Other Essays. Boston: Methuen.

Wallace, R. Jay. 2008. "Emotions, Expectations, and Responsibility," in Free Will and Reactive Attitudes: Perspectives on P. F. Strawson's "Freedom and Resentment". Ed. Michael McKenna and Paul Russell. Surrey, UK and Burlington, VT: Ashgate Publishing Company.

Other People

Kieran Setiya

(Draft; do not cite without permission)

Do you believe in love at first sight? Maybe you do and maybe you don't. Perhaps you will refuse to say, complaining that the question is obscure. I sympathize with that response. In a way, it is the subject of this essay, though I hope to show that there is more at stake. I begin with the prediction that, whatever you make of love at first sight, you do not believe in "love at definite description." You may know on general grounds that there is a shortest spy, but you cannot love the shortest spy if you have not met her and know nothing more about her.1 You could, I suppose, become invested in the prospects of the shortest spy, whoever she is, preferring outcomes that will benefit her to ones that benefit other people, striving to ensure that the shortest spy survives and flourishes. But this would not be love, and absent further context, it would not be rational.

It might be different if the description were more poignant: "the woman who saved your life" or "the brother you never knew." Special concern for individuals so described may be intelligible. Likewise, perhaps, if the description evokes, in richly textured detail, an attractive human being.

Personal acquaintance may be not be required for love. But mere description, as in "the shortest spy," is not enough. "Personal acquaintance," here, is a place-holder for the relation to another human being that justifies love at first sight, if there is any such thing. Personal acquaintance is the minimal cognitive contact that makes sense of love.[197] This paper explores the nature of this relation and its place in moral philosophy. As I will argue, personal acquaintance plays a role not just in love but in concern for individuals, as such.

Section I is about the connections between personal acquaintance, love, and moral standing. It maps some puzzling features of personal acquaintance that set parameters for any attempt to comprehend it. The task is to account for the ethical significance of this relation. In section II, we find a similar structure in concern for others of the sort that is morally required. This structure comes out in recent treatments of contractualism, aggregation, and the trolley problem. Section III turns to the work of Emmanuel Levinas as a source of insight into personal acquaintance, tracing the difficulties with his view and the prospects for revision. We are left with a question not just about love but about the basis of human values and the value of human life.

I

In "Love and the Value of a Life," I argued that it is rational for any one of us to love any other human being, whatever their merits, without the need for any past relationship (Setiya 2014: §1). In rejecting the need for virtues or common histories as grounds for love, I agree with David Velleman. For Velleman (1999: 366), "respect and love [are] the required minimum and optional maximum responses to one and the same value," the value of our rational nature. Like Velleman, I believe that the subjects of full moral

[197] The example derives from Kaplan 1968: 192-3.

standing, who deserve respect, coincide with those it is rational to love in the distinctive way that we love other people. By "full moral standing," I mean the kind of significance shared by human beings but not by other animals, at least not the sort we encounter on Earth. Our interests count for more than theirs, and we have rights against each another they do not possess. (We will come back to this assumption at the end.)

I differ from Velleman on three counts. First, I do not share his Kantian conception of the basis of moral standing, on which it turns on our rational nature. In my view, human beings who lack reason, and the potential for it, are morally equal to us. Second, I am less resistant than Velleman to the idea that, in its primary forms, love involves a disproportionate concern for the 3 interests of the beloved, concern that goes beyond what is required by moral standing.[198] While there are different varieties of love – erotic, parental, and so on – this a defining feature of the sort of love that interests me. Finally, while I doubt the need for past relationships as reasons for love, I do not deny that friendship, parenthood, and other relationships provide such reasons.[199] Parents should love their children, and friends should care for one another, albeit in different ways.

I won't argue for any of this now. Instead, I will concede an omission, brought out by love at definite description.[200] Even on the most permissive view of love, on which it does not turn on particular merits or past relationships, you cannot love the shortest spy on the basis of that description. What is possible, and rational, is love at first sight. So the position must be qualified. On the most permissive plausible view, it is rational to love any human being with whom you are personally acquainted, not any human being, full stop. But then we have to ask: what is personal acquaintance and how does it justify love?

Both the interest and the enigma of personal acquaintance come into focus if I am right about the implications of the permissive view of love. The most dramatic consequence speaks to the moral significance of numbers. Consider a case in which you can save the lives of three strangers drowning over to the left or a single stranger, M, who is drowning on the right.[201] The circumstance is otherwise unexceptional. You have no special obligation to any given stranger, and their survival would have no unusual consequences, good or ill. On my view, it would be rational for you to love M, even though you have never met before. What form can this love take? Must it be romantic, passionate, or possessive? No. It would be rational to feel dispassionate love for M, in which one is more strongly moved by her needs than by those of other people. Acting on this disposition, you might rationally decide to save the life of M instead of saving three. In general, where the disposition to φ in light of certain beliefs is rational, and those beliefs are true, one has sufficient reason to φ. It follows that you have sufficient reason to save the life of a single drowning stranger when you could save more. We have reached a

[198] Compare Velleman 1999: 353, Setiya 2014: 252-4.

[199] Setiya 2014: 258-62, responding to Kolodny 2003.

[200] The omission is partial: I appeal to "singular thought" at several points (Setiya 2014: 260n21, 265-6). Velleman has urged that emotions such as love depend on "acquaintance-based thought" (2008: 269- 70), though he does not develop the point and it is in tension with his earlier remarks about the attachment of adopted children to birth parents they have never met (see Velleman 2008: 263-4).

[201] The case derives from Anscombe 1967: 17.

version of John Taurek's (1977) startling conclusion that, in cases of this kind, the numbers do not count: at the very least, they are not decisive.

I don't expect this thumbnail sketch to be convincing: more argument is required.[202] But it shows how doubts about aggregation flow from the permissive view of love, assuming love can take the form of a dispassionate but disproportionate concern for someone's needs. Now for the puzzle. When I first drew these connections, I did not stress the role of personal acquaintance. It may be rational to save M at the cost of three lives when you are confronted with M herself: when you look into her eyes and feel the weight of her needs, responding with love. That is contentious enough. I do not think it would be rational to save the person on the right when you know them only by that description. In what we may call the "anonymous" case, you have no contact with the drowning strangers. You are merely told what is happening and must decide where to send the rescue mission. It is irrational to give priority to the needs of one in the anonymous case. You are not in a position to love the person on the right. That takes personal acquaintance.

The nature of personal acquaintance matters, on the permissive view of love, not just because it makes love rational but because, if love can be both partial and dispassionate, it makes a difference to questions of life and death. This brings out a pivotal constraint on how we conceive the relation of personal acquaintance. When you stand in this relation to M, it is rational to save her life, moved by the urgency of her needs, instead of the lives of the other three. When you lack this relation to M, when you know her only as "the one who is drowning on the right," it is irrational to save her life. Personal acquaintance is ethically significant. At the same time, it is utterly minimal, requiring no history of interaction, as we know from love at first sight. What can this relation be?

We may turn for help to philosophical discussions of "knowing who": to be personally acquainted with M is to know who she is. But accounts of "knowing who" in the philosophy of language only compound the mystery. On the minimal view, to know who someone is to know an answer to the question "Who is …?" The answer need not even be a definite description. David Braun begins his essay in defence of this conception with the sentence "Hong Oak Yun is a person who is over three inches tall," adding boldly: "now you know who Hong Oak Yun is" (Braun 2006: 24). In whatever sense, if any, this is true, it is not one that matters to moral philosophy or makes love rational.

On a more orthodox view, to know who someone is to know a contextually relevant answer to the question "Who is …?" that takes the form of a definite description.[203] But this does not amount to progress. At best, it frames our problem: which answers to the question "Who is …?" are ethically relevant? What do you need to know about someone in order to be personally acquainted with them and why does it matter? In fact, the situation is worse. In love at first sight, you know very little about the person you love apart from their relation to you. Knowing that they are the person with these properties is like knowing that they are shortest spy. It does not count as knowing who they are in an ethically relevant sense. The most plausible candidates for a description that matters, morally speaking, will be ones that cite your relationship to them. Why not then conclude

[202] I provide it in Setiya 2014.

[203] This is a drastic simplification of the theory proposed in Boër and Lycan 1986.

that this relation matters, not the further relation involved in knowing about it? The appeal to "knowing who" is a distraction.

We may turn instead to "rigid designation": the idea of a concept that essentially denotes a particular individual. Is the problem with "loving" the shortest spy or the person who is drowning on the right that their identity is not involved in one's response? They are picked out by properties they could lack. According to Philip Pettit, "when an agent displays a commitment to a beloved by acting out of love, the reason that moves the agent has to be rigidly individualized in favour of the beloved. It has to be a reason in which the beloved figures as an essential component" (Pettit 1997: 158-9). But again, this is not the point. Love at definite description remains irrational, or impossible, when the description is rigidified. It makes no sense to love the actual shortest spy or the person who is actually drowning on the right, picked out in those terms. Nor is the shift to naming consequential. Being told that the shortest spy is Ortcutt, or the drowning woman Pat, does not change your situation.

Finally, we may turn to the history of "acquaintance" as a pivotal term in the philosophy of mind. For Russell (1910-11) and others, acquaintance with particulars is what makes them available as direct objects of thought. Russell's views about this topic evolved over time and they are subject to interpretive dispute, but in his earliest phase, he seems to have believed that we are acquainted only with sense-data, universals, and the self. That idea has not fared well and many are now sceptical of any role for acquaintance as a condition of "singular thought."[204] Those who are sympathetic to the idea agree that a paradigm of acquaintance is perceptual contact of the sort that sustains demonstrative reference.[205] That is certainly present in love at first sight or the case in which you see the drowning M; and it is absent when you think of the shortest spy or the person who is drowning on the right. But this suggestion is puzzling. Why should perceptual contact, past or present, have the significance that personal acquaintance does? Why should seeing someone, or having seen them in the past, make it rational to give priority to their needs, to save their life at the cost of three? Perceptual contact may correlates with what matters, a relation made possible by perception; it is hard to see why it should matter in itself.

Personal acquaintance is an ethically significant relation. It is personal acquaintance that explains why it is rational to save one stranger when you could save three; its absence explains why it is not rational to do so in the anonymous case. An account of personal acquaintance must accord with its significance, as appeal to perceptual contact does not.

I have been describing a view on which it is rational to love any human being with whom one is personally acquainted but irrational, perhaps impossible, to love someone at definite description, at least when the description is as minimal as "the shortest spy." I want now to suggest that personal acquaintance is not just necessary but sufficient for the rationality of love. How could it fail to suffice? The idea would have to be that rational love depends on further beliefs, beliefs about the object of love. On the permissive view, these cannot be beliefs about his or her specific merits or about your past relationship. Nothing like that is required. Nor can we plausibly appeal to beliefs about the relation of

[204] For a recent critique, see Hawthorne and Manley 2012: Ch. 3.
[205] See, for instance, Dickie 2015: Ch. 4.

personal acquaintance. As before, it is the relation that counts, not knowledge of it. Must you believe that the object of love is a "person" in the philosophers' sense, a rational subject? No: you can love human beings who lack reason or the potential for it. Must you believe that the object of love is another human being? While it is irrational to love what you know to be a goat in the way you might love another person, as in the play by Edward Albee (2003), I don't believe that love depends on conjectured species or form of life. That the man across the room is a human being, not a rational Martian, is too theoretical a ground for love at first sight. Finally, we can ask if you must believe that the object of love has full moral standing. There is a sense in which you treat them as if they do; but you need not have beliefs about how they should be treated in order to be rational in loving them.

In principle, there might be other beliefs that justify love, other properties to which we must appeal. There is room for a disjunctive view, on which various beliefs will do. It is not easy to exhaust the options. But if we already know that personal acquaintance matters, that it is morally significant, unlike mere perceptual contact or referring to someone as "the shortest spy," why keep looking? Why not conclude instead that, given its ethical weight, personal acquaintance is sufficient to justify love.[206] As its name suggests, personal acquaintance is a relation we can bear only to those it is rational to love in the way that we love other people, only to those who have full moral standing. You cannot be personally acquainted with a goat, though you might believe you are. It is not a belief about someone that makes them available for love but the relation of personal acquaintance. In Wittgenstein's words: "My attitude to him is an attitude towards a soul. I am not of the opinion that he has a soul" (Wittgenstein 1953: 178).

If this is right, personal acquaintance is ethically significant in two ways. First, because its absence in the anonymous case explains why you cannot save one instead of three; its presence explains why you can. Second, because it is a relation we can have only to those with full moral standing. Each mode of significance constrains what personal acquaintance can be.

II

Do these issues pertain only to curious views about the nature and justification of love? I don't believe they do. Personal acquaintance plays a tacit but essential role in recent debates about contractualism and social risk.[207]

The puzzle for contractualists comes out in the following cases, described by Johann Frick. In Mass Vaccination (Known Victims), a million children face certain death unless they are treated with a vaccine, administered to all. Vaccine A cures the fatal disease but will leave the children with a paralyzed limb. Vaccine B cures the fatal disease without paralysis but "because of a known particularity in their genotype, [it] is certain to be completely ineffective for 1,000 identified children" (Frick 2015: 183). These children

[206] A case of particular interest is self-love. Surely this does not depend on the belief that you are a person, or a human being, or have moral standing. Nor, as I have argued elsewhere (Setiya 2015), does it rest on beleiefs about who you are. Instead, it turns on personal acquaintance with yourself.

[207] Contributions include: Scanlon 1998: 208-9, Reibetanz 1998, Ashford 2003, Lenman 2008, Fried 2012, James 2012, Dougherty 2013, Kumar 2015, Frick 2015.

will die. For contractualists, an act is permissible only if it can be justified to each of those affected, in that it is licensed by a principle none of them could reasonably reject. We are not allowed to aggregate claims. Thus, in Mass Vaccination (Known Victims), we compare the harm of losing one's life to the harm of a paralyzed limb. Since no-one can be asked to bear the former in order to save someone from the latter, we must choose Vaccine A.

Now consider Mass Vaccination (Unknown Victims). Here a million children face certain death unless they are treated with a vaccine. Vaccine A is available, but there is also Vaccine C, which cures the fatal disease without paralysis in 99.9% of cases; in o.1% of cases, it is utterly ineffective. (The probabilities here are epistemic: they reflect our evidence in making the decision.) The challenge for contractualism is to distinguish the second case from the first, given that the outcome of choosing Vaccine C is virtually certain to involve the death of at least one child, and very likely to involve the death of about 1,000.[208] According to Frick:

> [In] real life, we often impose social risks that closely resemble that of choosing [Vaccine C] in Mass Vaccination (Unknown Victims). Thus, it is commonly deemed morally unproblematic to systematically inoculate young children against certain serious but nonfatal childhood diseases where there is a remote chance of fatal side effects from the inoculation itself. (Frick 2015: 185)

Can contractualists explain why it is permissible to impose this kind of social risk while maintaining that it is impermissible to do so when the victims are identified in advance?

Frick's solution takes the form of "ex ante contractualism," according to which we should evaluate Mass Vaccination (Unknown Victims) not by considering how individuals fare in the possible outcomes but by considering how our policies affect their prospects now (Frick 2015: 187-8). The claim is that Vaccine C improves the ex ante prospects of each individual child, by our evidential lights. It gives them a 99.9% chance of total cure with a 0.1% chance of failure, which is arguably better than the assurance of paralysis with Vaccine A. That is how a policy of using Vaccine C can be justified to all. (If you believe that the imposition of a 0.1% chance of death on a given individual cannot be justified as the alternative to paralysis, reduce the risk until you agree. The general point remains.)

Ex ante contractualists thus permit the imposition of social risk while resisting the imposition of harms when the victims are known, or knowable, in advance.[209] It is important to stress that the dividing factor is not the chanciness of Vaccine C or the possibility that no-one dies. It is about identification. Consider a third case, Mass Vaccination (Unknown but Definite Victims), which is just like Mass Vaccination (Known Victims) except that there is no way to guess who has the distinctive genotype. Vaccine A will cure the fatal disease but leave the children with a paralyzed limb. Vaccine B will cure the disease without paralysis except for 1,000 unidentified children. For the ex ante contractualist, this case is like Mass Vaccination (Unknown Victims): Vaccine B improves the prospects of each child, by our evidential lights. No individual

[208] The likelihood is >0.99 that 1,000 children ±100 will die (Frick 2015: 183n14).

[209] On the extension from known to knowable victims, see Frick 2015: 191-3. I return to this below.

should object to our choosing Vaccine B even though, as in Mass Vaccination (Known Victims), 1,000 children are sure to die.[210]

Some will resist this verdict, assimilating victims who are definite but unknown to those who are known in advance. They will need to square their resistance with a plausible view of social risk. Why refuse to employ Vaccine B in Mass Vaccination (Unknown but Definite Victims) when it improves the prospects of each individual as much as Vaccine C? I won't pursue that issue here. Instead, I want to trace the implications of ex ante contractualism, drawing out an ethical idea that turns on personal acquaintance. In doing so, I will assume, for the sake of argument, that Frick's analysis is right.

The basic question for ex ante contractualists is what distinguishes Mass Vaccination (Known Victims) from Mass Vaccination (Unknown but Definite Victims), given that the objective probabilities of the various outcomes are the same. The terminology tells us that the difference is whether the victims are identified or known. But what exactly does that mean? It had better not suffice for a victim to be identified that we can pick them out by definite description. After all, we could "identify" the unknown victims by some irrelevant feature, like height: "the shortest child who has the gene"; "the second shortest child who has the gene"; and so on. We know that these children will not be saved by Vaccine B in Mass Vaccination (Unknown but Definite Victims). If that makes them "known victims," the alleged distinction will collapse. Suppose instead that we are given a list of names: these are the children who have the distinctive gene. We have no other way to determine who they are. Again, this seems ethically irrelevant. We knew all along that the children had names; it doesn't matter what they are. In contrast, I would urge, personal acquaintance must suffice for a victim to be identified or known, to transform the circumstance into Mass Vaccination (Known Victims), and so preclude the use of Vaccine B. What guides the ex ante contractualist is the idea of "personal concern": a concern for others directed at them as individuals, made possible, and rational, by personal acquaintance.

This leaves some difficult questions. Presumably, it is not required that we in fact identify or know the victims. For the ex ante contractualist, the question is what personal concern would motivate if we were personally acquainted with those involved, given what we know, or what is knowable, about them. In Mass Vaccination (Known Victims), concern of this kind does not speak with a single voice; for those who have the gene, it favours Vaccine A; for those who do not, Vaccine B. Where the victims are unknown, personal concern is arguably unanimous: it favours Vaccine B on behalf of each. That is why it is permissible to choose Vaccine B.

The idea, then, is not that you should be more concerned with personal acquaintances than anyone else, or that it is rational to give their interests greater weight. The idea is that, when you aim to justify a policy to each of those affected, their prospects on your evidence will depend on how you pick them out. In Mass Vaccination (Unknown but Definite Victims), the prospects of 12 the shortest child with the gene are very bad if she is given Vaccine B. But if you meet a random child, her prospects on your evidence look

[210] Unfortunately, Frick does not discuss this case, but he considers a variant of Mass Vaccination (Known Victims) in which the genetic test is very costly, and concludes, on ex ante contractualist lines, that it is permissible to choose Vaccine B; see Frick 2015: 193–4.

better with Vaccine B than Vaccine A. For the ex ante contractualist, the first way of picking children out, by definite description, is irrelevant: that is not how you should think of individuals when you ask whether a policy can be justified to each. In contrast, the second way of picking children out, by personal acquaintance, is morally apt.

Whatever you make of contractualism as a theory of right and wrong, the idea of personal concern, concern that is mediated by personal acquaintance, is ethically compelling. It is like love, as described in section I, except that it is not disproportionate, and like respect but unlike love, it is a response to others we are required to have. It is a form of impartial concern for individuals that personal acquaintance demands. Arguably, such concern is akin to love in that its justification does not turn on further beliefs about the object of concern. Personal acquaintance is again significant in two ways. First, because it justifies a kind of concern that has ethical weight in decisions that benefit others. Second, because it is a relation we can have only to those with full moral standing. Each mode of significance constrains what personal acquaintance can be.

I have argued that ex ante contractualists share the puzzle of personal acquaintance: the task of explaining its character in a way that meets these ethical constraints. But the idea of personal concern appears elsewhere. Perhaps the most self-conscious invocation of personal concern in recent moral philosophy is due to Caspar Hare (2016: §3). Hare begins with the standard Footbridge case, introduced by Judith Thomson (1976): you can push a button to drop one person from a bridge into the path of a speeding trolley that will otherwise kill five. Most believe that doing so would be wrong. Hare contrasts the original case with what we can call "Opaque Footbridge": six people you know and care about are caught up in the trolley case, five on the track, one on the bridge, but you do not and cannot know where in particular they are. As Hare contends, there is a powerful argument that concern for each of those involved counts in favour of pushing the button. If we give them alphabetical names, we can see that, by your lights, pushing the button will improve A's prospects from a 5/6 chance of death to just 1/6. It is true that pushing the button will change the potential cause of death, from being hit by a runaway 13 trolley to falling from a bridge as a result of your intervention. But from A's perspective, why care? Why should it matter whether you die on the tracks or falling from a bridge to save the five? The upshot is that, in Opaque Footbridge, concern for A alone, not weighing her interests against those of others or aggregating claims, should lead you to push the button. The same is true of concern for B, and C, and all the rest. Benevolence speaks with a single voice.

As Hare insists, this argument does not apply in the original Footbridge case (Hare 2016: 466). Again, suppose you know the six involved, from A to F. If you know that F is on the bridge, concern for each is not unanimous. There is no way to argue that you ought to push the button without comparing or combining claims. Benevolent concern is simply divided. Concern for F speaks against pushing the button, concern for the others speak in favour. This conflict cannot be ignored.

Hare gives further arguments.[211] But we need not go into them. Nor need we accept his conclusion that, in Opaque Footbridge, you ought to push the button.[212] What matters is that, regardless of this conclusion, Hare's argument taps an ethical idea that has real force. He seems right to insist that in Opaque Footbridge, concern for the interests of those involved speaks unanimously for pushing the button. If there is a moral objection to doing so, it does not flow from benevolent concern but from a different and potentially conflicting source: a respect for rights that is not grounded in and may diverge from people's interests.

As with ex ante contractualism, this reasoning appeals to personal concern: concern for individuals that rests on personal acquaintance. We can see this by asking what explains the contrast between Footbridge and Opaque Footbridge. The answer is that, in Opaque Footbridge, you do not know who will die if you push the button, whereas in Footbridge, you do: the victim is identified or known. As before, it had better not suffice for identification that you locate someone by description, since you can "identify" the victim in Opaque Footbridge as "the one who is on the bridge." If that makes them an identified victim, the contrast we are tracking disappears. Nor do names seem ethically relevant. Nothing changes when you are told that the person on the bridge is "Jim" – unless you know him in some other way.

In what meaningful sense, then, do you know who the victim is in Footbridge but not in its opaque counterpart?[213] Confronted with this question, Hare contends that the sort of "knowing who" that makes a difference is knowing facts about what matters in the lives of those involved, about their friends and families, hobbies and careers. What blocks the argument for pushing the button is the plurality of values realized by the diverse activities of A to F: these values are incommensurable. What blocks the argument for pushing the button is the plurality of values realized by these diverse activities: values that are incommensurable (Hare 2016: §6). But this can't be the right account. It would not affect the ethics of Footbridge if the people involved were perfect duplicates of one another, identical sextuplets who lead identical, solitary lives. Nor would it matter if they were people you just met, about whom you know nothing at all. What counts is personal acquaintance, not biographical knowledge. In Footbridge, personal concern for the one who is on the bridge restrains you from pushing the button. In Opaque Footbridge, personal concern – concern for individuals that turns on personal acquaintance – speaks in favour. Concern for the person on the bridge, described as such, can be ignored.

Again, the moral of the story is that personal concern has ethical weight. It is not that you should be more concerned with personal acquaintances than anyone else, or that it is rational to give their interests greater weight. The idea is rather that, when you care about people's interests, their prospects on your evidence depend on how you pick them out. In Opaque Footbridge, the prospects of the person on the bridge are bleak if you push the button. But the prospects of A to F, picked out by personal acquaintance, all improve. It

[211] His strategy is to decompose your action into six, each of which affects only one individual, improving their prospects without affecting anyone else. For details, see Hare 2016: §4.
[212] I object to it in Setiya ms.
[213] A question raised about a similar case by Elizabeth Harman, in her review of Hare 2013; see Harman 2015: 870.

is the second fact that bears on concern for the interests of those involved. In order to make sense of this, to see the contrast between Footbridge and Opaque Footbridge, we must appeal to a form of concern that attaches to 17 A question raised about a similar case by Elizabeth Harman, in her review of Hare 2013; see Harman 2015: 870. 15 individuals not by name or description but by personal acquaintance. Such concern resembles love, except that it is not disproportionate and is not merely rational but required. It is tempting to add, once more, that the justification for personal concern does not depend upon beliefs about its object: personal acquaintance is enough. It is a relation we can have only to those with full moral standing.

There are thus three routes to the puzzle of personal acquaintance. It follows from the permissive view of love, from ex ante contractualism, and from Hare's appeal to concern for others in Opaque Footbridge, that personal acquaintance justifies a kind of concern that makes a difference, either to saving one or three, to the imposition of social risk, or to the reasons for pushing the button.[214] My hope is that, even if you doubt the premise of each argument, you can feel the pull of personal concern as an ethical idea. Non-aggregative, distributed concern for individuals with whom one is personally acquainted: this makes moral sense. Concern that is mediated by definite descriptions or the second-hand use of names does not. An account of personal acquaintance should explain why. III What is personal acquaintance? It is a relation that justifies both love and personal concern. In the first case, one's response is selective and disproportionate. In the second case, it is not. Instead, it is a kind of concern we should invest in everyone, a concern that is mediated by personal acquaintance. We should care about the interests of others as if were personally acquainted with them. When we weigh the effects of our actions on the prospects of individuals, it matters how we pick those individuals out. A's prospects on our evidence may differ from the prospects of the person on the bridge when, unbeknownst to us, A is the person on the bridge. Which way of picking people out is morally relevant? It is the one involved in personal concern, which attaches to individuals by way of personal acquaintance. Personal acquaintance plays a role in determining the object of our attitude that is elsewhere played by definite descriptions or the second-hand use of names. This mode of presentation is deployed in thoughts – for instance, beliefs about the prospects of a given individual – that interact with such concern. In Fregean terms, personal acquaintance is the basis of distinctive singular concepts; alternatively, it is a guise under which we can think of others. Either way, it is a cognitive relation that individuates its object, sustaining reference, and it is a relation that has moral weight. What more can we say?

In the work of Emmanuel Levinas, spanning four decades of the mid-twentieth century, we read what I think is a profound phenomenology of personal acquaintance.[215] Levinas comes back again and again to the face of the other as an ethical address. This theme is central to his most well-known book, Totality and Infinity (1961). But his argument is sketched in "Freedom and Command," published in 1953:

---

[214] As I argue in Setiya ms., there is a fourth route, too, through the nature of respect for rights.

[215] I am no expert on Levinas, but I have been inspired by his writings. Michael Morgan's (2007) Discovering Levinas is an invaluable guide; I have also been helped by Perpich 2008.

> The being that expresses itself, that faces me, says no to me by this very expression. This no is not merely formal, but it is not the no of a hostile force or a threat; it is the impossibility of killing him who presents that face; it is the possibility of encountering a being through an interdiction. The face is the fact that a being affects us not in the indicative, but in the imperative, and is thus outside all categories. […] The metaphysical relationship, the relationship with the exterior, is only possible as an ethical relationship. (Levinas 1953: 21)

Levinas is as much concerned with justice ("That shalt not kill") as with benevolence, though he connects the two:

> From the start, the encounter with the Other is my responsibility for him. That is the responsibility for my neighbor, which is, no doubt, the harsh name for what we call love of one's neighbor; love without Eros, charity, love in which the ethical aspect dominates the passionate aspect, love without concupiscence. (Levinas 1982a: 103)

Levinas insists on the particularity of our relation to the other, its distributed, non-aggregative character, in ways that resonate with personal concern.

> I must judge, where before I was to assume responsibilities. Here is the birth of the theoretical; here the concern for justice is born, which is the basis of the theoretical. But it is always starting out from the Face, from the responsibility for the other that justice appears, which calls for judgment and comparison, a comparison of what is in principle incomparable, for every being is unique; every other is unique. (Levinas 1982a: 104)[216]

For Levinas, our relation to the other, face-to-face, is always already ethical: it affects us in the imperative, not the indicative. He does not try to justify this relation, or explain its basis in other terms. To many philosophers, this will seem like an abdication of responsibility. What grounds the ethical phenomena Levinas describes? What cognitive relation justifies love at first sight and mediates personal concern, a form of concern than structures ethical thought? Since the ethical supervenes on the non-ethical, there must be an answer to this question.[217] Isn't that where personal acquaintance comes in? As I read him, however, Levinas does not believe that the gap can be filled, that personal acquaintance can be specified except in ethical terms, as the relation that plays these roles.[218] And it's difficult to see how it could be done.

What psychological description can we give of this relation? What can we add to perceptual contact to explain why personal acquaintance matters and how it implicates moral standing? One idea is to look at the facts to which we gain perceptual access. Personal acquaintance might involve perceptual contact of a kind that affords perceptual knowledge of properties that matter, morally speaking. For instance, it might allow for knowledge of mental states. When we are personally acquainted with someone, the suggestion runs, we can perceive their joy and suffering, weal and woe. Whether or not that is true, however, it is doubly unpromising. First, it gets the extension wrong. If we can perceive human suffering, why not the suffering of non-human animals, who lack

[216] On the particularity of ethics in Levinas, see Morgan 2007: 61, 79-80.
[217] I discuss supervenience in Setiya 2012: 8-11.
[218] Here I follow Morgan 2007: 46-50; see also Perpich 2008: 51-4, 74-5, 115-7.

moral standing of the sort at issue here? Second, it is hard to see why the perception of suffering, or its possibility, should matter more than knowledge of human suffering acquired by other means. Why would it justify a distinctively personal concern? The second problem applies to variations of this approach that turn on perceptual access to specifically human qualities, to perception of the face, or mind, or body, that brings it under concepts specific to human life. Views of this kind fare better extensionally, but they do not accord with the moral weight of personal acquaintance. If it is simply a matter of how we know about the other, personal acquaintance should not matter in the ways it does on the views discussed above. For Levinas, "[the] encounter with the face is not an act of seeing; it is not perceptual or judgmental" (Morgan 2007: 75).[219]

What goes missing in the turn to perceptual knowledge is the practical dimension of personal acquaintance. One way to fill this deficit is to the stress the role of perceptual contact as a basis for human interaction. Personal acquaintance matters, on this more Kantian approach, because it allows us to act and reason together. For Christine Korsgaard, "the violation of a deontological constraint always involves an agent and a victim, and thus […] deontological reasons are always shared reasons. They cannot be the personal property of individual agents. Instead, they supervene on the relationships of people who interact with one another. They are intersubjective reasons" (Korsgaard 1993: 298). That might explain why personal acquaintance counts. It is in the spirit of Stephen Darwall's (2006) invocation of the "second-person standpoint," the point of view from which we make claims on one another, holding each other accountable, you and I.

Is reciprocal recognition, or a nexus of rational wills, the ground of personal concern? I don't believe it is. The view in question could take various forms but they share two basic flaws. The more mundane objection is again extensional. Human beings with whom we cannot interact as agents have full moral standing. They are rational objects of love and personal concern. This is true even if they lack the potential to interact with us. I don't know how to prove that infants with irreparable cognitive disabilities and people in persistent vegetative states are morally equal to us, and I do not think the implications of this fact are clear, but I am quite sure that it is true.

The less mundane objection is phenomenological. Though Darwall cites both Levinas and Martin Buber (1923) as precedents for the second-person standpoint, their views are not the same.[220] Buber appeals to the reciprocity of the "I-Thou" relation. Levinas emphatically does not.

> [The] relationship with the other is not symmetrical, it is not at all as in Martin Buber. When I say Thou to an I, to a me, according to Buber I would always have that me before me as the one who says Thou to me. Consequently, there would be a reciprocal relationship. According to my analysis, on the other hand, in the relation to the Face, it is asymmetry that is affirmed: at the outset I hardly care what the other is with respect to me, that is his own business; for me, he is above all the one I am responsible for. (Levinas 1982a: 105)

[219] See also Morgan 2007: 92.

[220] On Levinas, see Darwall 2006: 21-22n44; on Buber, Darwall 2006: 39-40.

> One of the themes of Totality and Infinity […] is that the intersubjective relationship is a non-symmetrical relationship. In this sense, I am responsible for the other without waiting for reciprocity, were I to die for it. Reciprocity is his affair. (Levinas 1982b: 98)[221]

On this point, I think Levinas is right. The phenomenology of personal acquaintance is not mutual or interactive: the demand for personal concern is unilateral. It is about what I owe to you not what we owe to each other. This ethical reality is obscured by the Kantian focus on the second person. We should not conflate attention to relational phenomena in ethics – not just personal concern but the relational or bipolar notion of wronging an individual – with appeal to reciprocal recognition.[222]

Though it is impossible to survey every possibility, I hope you can begin to see how hard it is to state the nature of personal acquaintance in psychological terms: to identify a psychological relation we can bear only to those with full moral standing, a relation that justifies love and necessitates personal concern. It is no accident that Levinas does not describe the basis of the face-to-face relation. There is an echo of Wittgenstein in this refusal: "If I have exhausted the justifications I have reached bedrock, and my spade is turned" (Wittgenstein 1953: §217). Cora Diamond takes a similar view of membership in the moral community:

> The sense of mystery surrounding our lives, the feeling of solidarity in mysterious origin and uncertain fate: this binds us to each other, and the binding meant includes the dead and the unborn, and those who bear on their faces 'a look of blank idiocy,' those who lack all power of speech, those behind whose vacant eyes there lurks 'a soul in mute eclipse'. I am not arguing that we have a moral obligation to feel a sense of solidarity with all other human beings because of some natural or supernatural property or group of properties which we all have, contingently or necessarily. I am arguing, though, that there is no need to find such a ground… (Diamond 1991: 55)

Levinas in fact goes further. My relation to the other not only lacks a rationalpsychological ground, it is primitive or irreducible. There is no way to articulate its content in other terms. This relation is ethical through and through. At the same time, it is a condition of meaningful communication, which is a condition of public language, which is a condition of rational thought. (Like many philosophers, Levinas sees a distinction of kind between our mental lives and the "non-conceptual" psychology of non-linguistic animals.) Our ethical relation to the other is therefore presupposed by openness to the world: "the order of meaning, which seems to me primary, is precisely what comes to us from the interhuman relationship, so that the Face, with all its meaningfulness as brought out by analysis, is the beginning of intelligibility" (Levinas

---

[221] See also Morgan 2007: 62

[222] This distortion affects even those who resist the Kantian line. In a broadly Aristotelian approach to bipolarity, Michael Thompson assumes that "relations of right" are fundamentally reciprocal: in the paradigm case, they are recognized on both sides, though there may be marginal occaions in which the party who is wronged is unable to recognize the obligation of the other (Thompson 2004: 348, 367- 72). If I understand him, Levinas would question this assumption.

1982a: 103). The ethics of the face-to-face, of love and personal concern, is the transcendental origin of thought, as such.[223]

What I have just portrayed in outline is a central argument of Totality and Infinity.[224] It is "transcendental" in the Kantian sense: it aims to undermine a sceptical threat by showing how the sceptic's position assumes or implies the very thing she purports to doubt. In this case, the moral sceptic cannot think conceptually without relying on a public language that depends in turn on her ethical acknowledgement of the other. For Levinas, "[to] kill is not to dominate but to annihilate; it is to renounce comprehension absolutely" (Levinas 1961: 198).

I mention this argument not to endorse it but to give a more adequate picture of Levinas on the ethical roots of metaphysics, and to explain how the ineluctably ethical character of personal acquaintance or the face-to-face might bear on moral philosophy. Those are topics to pursue elsewhere. I want to return, instead, to the supervenience of the ethical: the pressure to insist that the justification of love and personal concern derives from a relation to the other we can specify in other terms. As we have seen, it is difficult to meet this pressure, to give a psychological account of personal acquaintance. What psychological relation makes love rational and calls for personal concern? The relations we have considered are extensionally wrong or ethically insignificant or both. Must we concede that, in this respect, morality is groundless?

Perhaps there is another way. Suppose, to begin with, that love and personal concern are natural kinds, emotions that play particular, distinctive roles in human life. Suppose, further, that there are regulated by a relation, R, that can be specific in psychological terms. And adopt the conjecture that R is personal acquaintance. Human beings feel love or personal concern for those with whom they are personally acquainted, not those who are known to them merely by name or by minimal description, like "the one on the left" or "the shortest spy." We should treat this as a generic proposition, a claim about what is characteristic of us that allows for exceptional cases, in which our emotions are misdirected. The psychological relation we are targeting is one by which they are naturally regulated, though the regulation may be imperfect. Suppose, finally, that the psychological relation thus described is one that relates human beings only to those with full moral standing: presumably, in the first instance, other human beings. We cannot be personally acquainted with inanimate objects or with non-human animals of the sort we encounter on Earth.

The discussion so far has asked, in effect, why relation R would justify love and necessitate personal concern. It treats our hypothesized emotions as if they were in need of external vindication, holding human nature up to a normative standard independent of us. Could that be a mistake? What if we insist that human nature, and the facts of human

[223] "Preexisting the disclosure of being in general taken as basis of knowledge and as meaning of being is the relation with the existent that expresses himself; preexisting the plane of ontology is the ethical plane" (Levinas 1961: 201).

[224] See, especially, Levinas 1961: 72-81, 194-219. This argument is explored by Morgan (2007: 52-5) and Perpich (132-5, 140-9). An early version appears in Levinas 1953: 18.

life, play a constitutive role in ethics, as in the tradition that descends from Aristotle?[225] That a human response is rational or justified is not independent of the fact that this response, or affirmation of this response, is functional for us. We need not read the virtues directly or naïvely from the book of human life in order to accept some measure of constitutive dependence. In fact, we had better not, unless we believe that human beings are by nature perfectly good. The devil is in the details.[226] But the approach has merit, in part because it is the only way we have seen, thus far, to reconcile the ethics of personal acquaintance with its psychological grounds. On this view, personal acquaintance matters not because it ought to play a certain role in human life but because it does: it is the relation that underlies both love and personal concern. Personal concern is called for, and love is justified, whenever they are humanly possible.

There is more to say in defense of these ideas. Because I don't know how to say it, I want to end, instead, by placing the puzzle of personal acquaintance in a wider context of reflection on human values. At the beginning of section I, I assumed without argument that human beings have an ethical significance that is not shared by other terrestrial animals. Our interests count for more than theirs, and we have rights against each other they do not possess. Positions of this sort have acquired a very bad name. Do they reflect an odious form of "speciesism"?[227] It helps to emphasize their relational character: they are about the significance we have for one another, not about the significance of human beings in some absolute sense, as though we should matter more to rational Martians than they do to themselves. But even with this proviso, the basic challenge remains. How is such "humanism" (as I prefer) morally better than racism or sexism, attributing ethical significance to brute biological difference?[228] This question, which casts doubt on the distinctive value of humanity, has less force if human nature is involved in the foundations of ethics. If human beings by nature respond to one another in distinctive ways, as with love or personal concern, and this fact plays a constitutive role in how it is rational to respond, humanism might be true. By contrast, there is no credible theory of ethics on which its foundations appeal to race or sex, nor is there reason to believe that human beings are by nature racist or sexist in ways that might support an Aristotelian defence of such repugnant views.[229] There is, if not a direct argument from humanism to Aristotelian ethics, at least an affiliation between the two.

The ethics of personal acquaintance amplifies and complicates this connection. It is, to begin with, another instance of moral thinking that is difficult to sustain if we deny a constitutive role in ethics to the facts of human life. Perhaps we should not hope to

[225] See Foot 2001; Thompson 2013.

[226] I defend a qualified view, which finds a place for human nature as the ground of ethical knowledge, in Setiya 2012: Ch. 4.

[227] The term was coined by Peter Singer (1975: 6).

[228] This challenge is cenral to Singer's (1975) argument; for a more recent discussion, see McMahan 2005: §3.

[229] I discuss this point in Setiya 2014: 142-58.

sustain these thoughts, but if we do, we will be led, through Levinas, to Aristotle. At the same time, personal acquaintance puts constraints on the nature of moral standing: it has to mesh with human psychology in ways hypothesized above.

This points to a final question, often raised as an objection to humanism: what about rational Martians? Don't members of other rational species count for us in the same way other humans do? The standard response, which I accept, is that humanism does not imply otherwise. What it suggests is not that rational Martians lack full moral standing but that, if they have it, the ground on which they do so is quite different from the ground that applies to you or me. Whether we should care about the members of another rational species, what rights they have against us: these are open questions. The answers turn on how they relate to one another and to us. (Bernard Williams makes this vivid by imagining rational predators who come from outer space.[230] We can also conceive of rational beings who regard one another as prey.)

The idea of personal acquaintance introduces something new. For there is nothing in the psychology of love or personal concern that prevents us from being personally acquainted with non-human beings. One thing we learn from unimaginative science fiction, in which the aliens are mostly humanoid, is that love across species boundaries makes sense. The same is true of personal concern. If it is rational to love the members of another rational species, their moral standing should not be in doubt. The ethics of personal acquaintance is not humanist in giving special weight to specifically human life. It is humanist in treating every human being as a moral equal and, in its Aristotelian form, in giving special weight to human values, values that may be cosmically cosmopolitan.

We have traveled far along a speculative path. Let us go back to the start. I have argued that personal acquaintance plays a crucial role in the permissive view of love, in ex ante contractualism, and in Hare's account of Opaque Footbridge. If we want to make sense of these phenomena, we need an ethics of personal acquaintance. But it is hard to say what personal acquaintance is in terms that would explain why it justifies love and calls for personal concern. We have considered an approach that has some promise, one that draws on Aristotle's ethics, echoing Levinas on the face-to-face relation without his quietism. It is worth pursuing, though I am not sure that it works. If it doesn't, we are left with a serious, unsolved puzzle. Can we make sense of love at first sight, and of concern for individuals, as such?[231]

References

Albee, E. 2003. The Goat, or Who is Sylvia? London: Penguin.

Anscombe, G. E. M. 1967. Who is wronged? Oxford Review 5: 16-17.

Ashford, E. 2003. The demandingness of Scanlon's contractualism. Ethics 113: 273-302.

[230] See Williams 2006: 149-52.

[231] For discussion of this material in earlier forms, I am grateful to Alex Byrne, Sarah Buss, Imogen Dickie, Jimmy Doyle, Marah Gubar, Caspar Hare, Samia Hesni, Abby Jaques, Ryan Preston-Roedder, Tamar Schapiro, Jack Spencer, Quinn White, and Steve Yablo, to audiences at Brown University, the Normativity Research Group in Montreal, and MIT.

Boër, S. E. and Lycan, W. G. 1986. Knowing Who. Cambridge, MA: MIT Press.

Braun, D. 2006. Now you know who Hong Oak Yun is. Philosophical Issues 16: 24-42.

Buber, M. 1923. I and Thou. Translated by W. Kaufmann. New York, NY: Touchstone, 1970.

Darwall, S. 2006. The Second-Person Standpoint. Cambridge, MA: Harvard University Press.

Diamond, C. 1991. The importance of being human. In D. Cockburn, ed., Human Beings. Cambridge: Cambridge University Press, 1991: 35-62.

Dickie, I. 2015. Fixing Reference. Oxford: Oxford University Press.

Dougherty, T. 2013. Aggregation, beneficence, and chance. Journal of Ethics and Social Philosophy 7: 1-19.

Foot, P. 2001. Natural Goodness. Oxford: Oxford University Press.

Frick, J. 2015. Contractualism and social risk. Philosophy and Public Affairs 43: 175-223.

Fried, B. 2012. Can contractualism save us from aggregation? Journal of Ethics 16: 39-66.

Hare, C. 2013. The Limits of Kindness. Oxford: Oxford University Press.

Hare, C. 2016. Should we wish well to all? Philosophical Review 125: 451-472.

Harman, E. 2015. Review of Caspar Hare, The Limits of Kindness. Ethics 125: 868-872.

Hawthorne, J. and Manley, D. 2012. The Reference Book. Oxford: Oxford University Press.

James, A. 2008. Contractualism's (not so) slippery slope. Legal Theory 18: 263-292.

Kaplan, D. 1968. Quantifying in. Synthese 19: 178-214.

Kolodny, N. 2003. Love as valuing a relationship. Philosophical Review 112: 135-89.

Korsgaard, C. 1993. The reasons we can share. Reprinted in Creating the Kingdom of Ends. Cambridge: Cambridge University Press, 1996: 275-310.

Kumar, R. 2015. Risking and wronging. Philosophy and Public Affairs 43: 27-49.

Lenman, J. 2008. Contractualism and risk imposition. Politics, Philosophy, and Economics 7: 99- 122. 27

Levinas, E. 1953. Freedom and command. Reprinted in Collected Philosophical Papers. Translated by A. Lingis. Pittsburgh, PA: Duquesne University Press, 1998: 15-23.

Levinas, E. 1961. Totality and Infinity. Translated by A. Lingis. Pittsburgh, PA: Duquesne University Press, 1969.

Levinas, E. 1982a. Philosophy, justice, and love. Reprinted in Entre Nous: On Thinking-of-the-Other. Translated by M. B. Smith and B. Harshav. New York, NY: Columbia University Press, 1998: 103-121.

Levinas, E. 1982b. Ethics and Infinity: Conversations with Philippe Nemo. Translated by R. A. Cohen. Pittsburgh, PA: Duquesne University Press, 1985.

McMahan, J. 2005. Our fellow creatures. Journal of Ethics 9: 353-380.

Morgan, M. 2007. Discovering Levinas. Cambridge: Cambridge University Press.

Perpich, D. 2008. The Ethics of Emmanuel Levinas. Stanford, CA: Stanford University Press.

Pettit, P. 1997. Love and its place in moral discourse. In R. Lamb, ed. Love Analyzed. Boulder, CO: Westview Press, 1997: 153-163.

Reibetanz, S. 1998. Contractualism and aggregation. Ethics 108: 296-311.

Russell, B. 1910-11. Knowledge by acquaintance and knowledge by description. Proceedings of the Aristotelian Society 11: 101-128.

Singer, P. 1975. Animal Liberation. New York: Random House.

Taurek, J. 1977. Should the numbers count? Philosophy and Public Affairs 6: 293-316.

Thompson, M. 2004. What is it to wrong someone? In R. J. Wallace, P. Pettit, S.

Scheffler and M. Smith, eds., Reason and Value. Oxford: Oxford University Press, 2004: 333-384.

Thompson, M. 2013. Forms of nature: ‘first’, ‘second’, ‘living’, ‘rational’, and ‘phronetic’. In G. Hindrichs and A. Honneth, eds., Freiheit Stuttgarter Hegel-Kongress 2011. Frankfurt: Klostermann, 2013: 701-735.

Thomson, J. J. 1976. Killing, letting die, and the trolley problem. The Monist 59: 204-217.

Setiya, K. 2012. Knowing Right From Wrong. Oxford: Oxford University Press.

Setiya, K. 2014. Love and the value of a life. Philosophical Review 123: 251-280.

Setiya, K. 2015. Selfish reasons. Ergo 2: 445-472.

Setiya, K. ms. Ignorance, beneficence, and rights.

Velleman, J. D. 1999. Love as a moral emotion. Ethics 109: 338-374.

Velleman, J. D. 2008. Persons in prospect. Philosophy and Public Affairs 36: 221-288.

Williams, B. 2006. The human prejudice. In Philosophy as a Humanistic Discipline. Princeton, NJ: Princeton University Press, 2006: 135-152.

CHICAGO ATTRACTIONS

John Hancock Tower:
The best-kept secret in Chicago tourism is the Signature Lounge, located on the 96th floor of the Hancock Tower, 875 N. Michigan Ave. This bar /restaurant provides guests with a 360 degree
view of Chicago and Lake Michigan for the price of a drink --there is no admission fee.

The Magnificent Mile:
Chosen as one of the ten great avenues of the world , the Mag Mile is located just north of the loop and is Chicago's most prestigious shopping district. Water Tower Place, a very large mall, is located at 835 N. Michigan Avenue. Walking south on Michigan Ave (or taking any of the many buses) you will end at the Wrigley Building down on the river (which you can follow into the loop and to Millenium Park and the Art Institute).

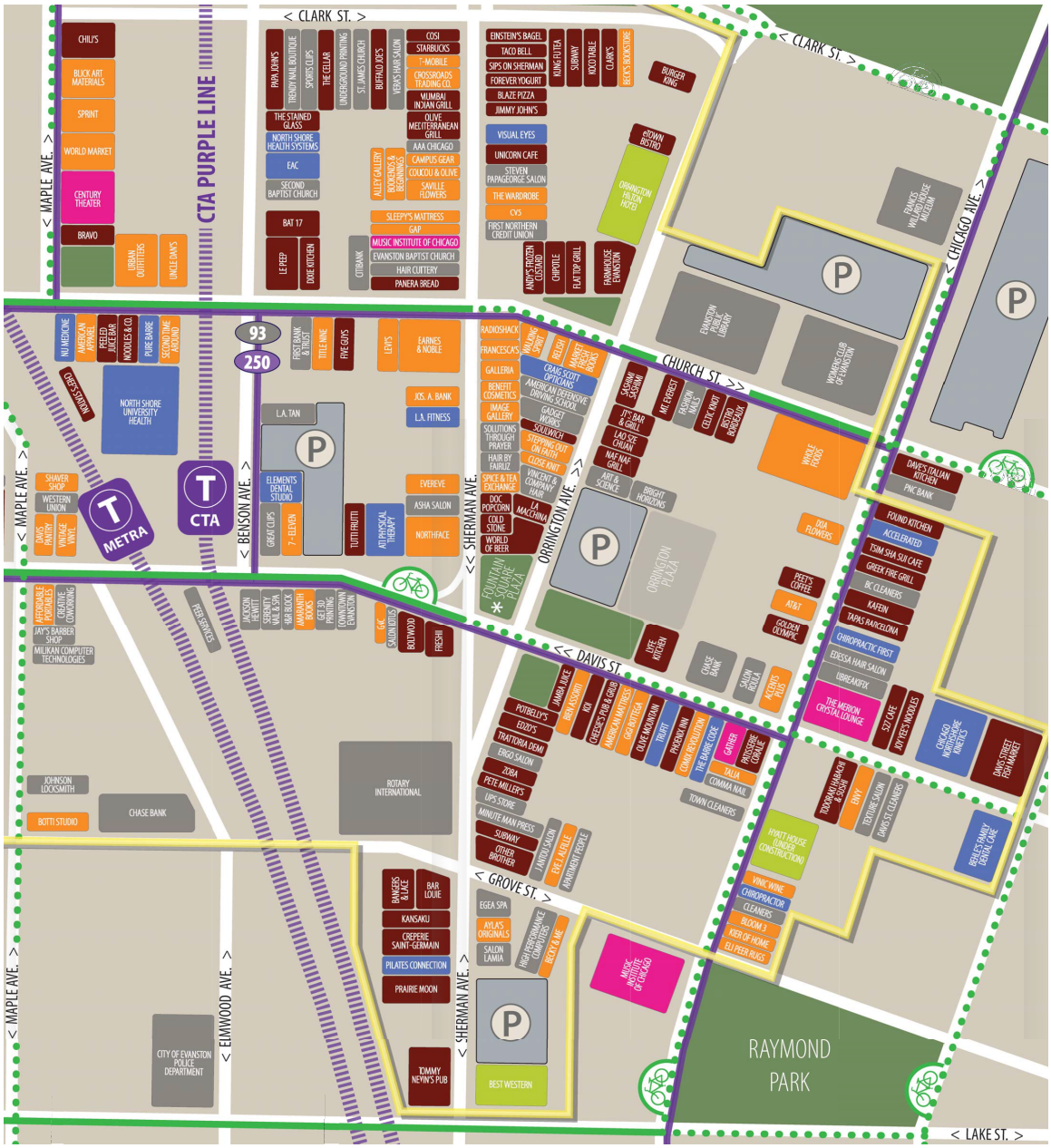
Chicago Architecture Foundation Boat Tour:
$44 for daytime crusies and $46 for nightime cruises, 90 minutes long. Dock location is southeast corner of the Michigan Avenue Bridge and Wacker Drive. Look for the blue awning marking the stairway entrance. You can buy tickets online.

Millennium Park:
Millennium Park is located in the heart of downtown Chicago. It is bordered by Michigan Avenue to the west, Columbus Drive to the east, Randolph Street to the north and Monroe Street to the south. This park is open daily from Sam to 11pm. Admission is free. Attractions include the enormous mirror-surfaced bean sculpture, the Cloud Gate bridge, the Crown Fountains, the outdoor ampitheater, and the Lurie Garden.

Shedd Aquarium:
Museum Hours: Weekdays: 9am-5pm & Weekends: 9am-6pm. Admission: $8 adults for aquarium only, $31 for all-access pass that includes Oceanarium, Wild Reef, Amazon Rising, the Caribbean Reef, Waters of the World, and others. To get to the museum, take the red line L to the Roosevelt stop and board a museum trolley or take the # 12 bus.

The Field Museum:
Museum Hours: 9am-5pm. $38 for an all-access pass. Take the red line L to the Roosevelt stop and board a museum trolley or take the # 12 bus.

Downtown Evanston



EAT/DRINK

SHOP

BE ENTERTAINED

STAY